Children's instrumental copying in play: comparing how children copy when they encounter evidence of failure in a close- and open-ended task

LAVENDER-FORSYTH, GUY,ALEXANDER

Children’s instrumental copying in play: comparing how children copy when they encounter evidence of failure in a close- and open-ended task

Guy A. Lavender Forsyth, BSc.

Supervised by Dr Rachel Kendal and Dr Jeremy Kendal

A thesis for the degree of MSc by Research in Evolutionary Anthropology,

Department of Anthropology,

Durham University

September 2018

Table of contents **Page no.**

List of tables **Page no.**

List of figures **Page no.**

Declaration

I confirm that no part of the material presented in this thesis has previously been submitted for a degree in this or any other institution. If material has been generated through joint work, this has been indicated where appropriate. All other sources have been referenced.

Acknowledgments

I am extremely grateful to my two supervisors, Rachel Kendal and Jeremy Kendal, for their unfailing support and guidance.

I also thank Zarja Muršič for her help in both this project's conception and execution.

I thank Robin Watson for his help in re-coding data for inter-observer analyses.

I also thank Dew Chutimon Kitviriya for her performance as the experimental model.

I am indebted to the workers at the Centre for Life for allowing me to collect data in their workplace.

I thank my parents for their interest, encouragement, and help.

Finally, I thank all of my participants, and their guardians, for their gracious involvement in the project.

Abstract

Recent research into play describes it as a context in which children are not constrained by goals external to their activities (i.e., their activities are open-ended), and in which children explore and learn about their environment. Yet research into children's social learning often underemphasises such open-ended contexts in favour of close-ended tasks in which children may or may not copy specific strategies to achieve pre-specified goals. I therefore examined two predictions, that: (1) children's copying behaviour will exhibit differences between close-ended and open-ended tasks, and (2) in the open-ended task, children will flexibly combine microstructural and macrostructural information learnt either 'socially' or 'asocially'. The second prediction did not find support in the data: data indicated that children given the open-ended task showed no differences between copying of microstructure design and copying of macrostructure design. However, the thesis presents results in support of the first prediction. Compared to the close-ended task, in which a successful rather than unsuccessful social model increased children's copying behaviour, data did not give reliable support for the same effects in the open-ended task. Counter to expectations drawn from literature using close-ended tasks, data from the open-ended task also gave no support to the hypothesis that children who encounter greater evidence of failure in their own building would rely more on social information. I suggest the explanation for these results is that incentives for instrumental copying are different when children are given a specific goal to achieve compared to when they have greater freedom to determine the aims of their activities. This study thus extends existing research by investigating an ecologically relevant, yet currently understudied, context for children's copying behaviour.

Chapter 1: Introduction

There is a common problem in the study of humans' evolution, which Frans de Waal describes in his 2016 best seller (*Are we smart enough to know how smart animals are?*): it is obvious that humans have a very different set of capabilities to other animals, and yet each of our species-wide capabilities, when examined closely, is arguably less exceptional than it appears. Two examples are the acquisition of skills such as language (often spoken, sometimes signed, and also written), and the use of increasingly complex and mechanically opaque tools. Many abilities required for language use, for example, are now widely thought to have evolutionary histories far deeper than *Homo* (Fitch 2017; Seyfarth & Cheney 2016). There is no single component of language cognition which researchers can agree makes the crucial difference to make human language different (e.g., Hauser et al. 2014). The same is true for the manipulation of ecological conditions through tool use (Kabadayi & Osvarth 2017; Vale et al. 2017a; Fausto & Valentina 2017), which is now observed in various forms across a large array of phylogenetically diverse animals (Botting, van der Waal & Rendell 2018). The solution to this problem of the source of the 'human difference' is at once obvious: there are no doubt countless small changes which humans accumulated over their evolutionary past, and which probably interacted with each other to produce our current state of affairs (see Sterelny 2012). This is an unsatisfying answer though; it amounts to stating that human evolution is complex, and we already suspected that. But embrace of the piecemeal approach to understanding human complexity (e.g., Stout & Hecht 2017) can create a starting point with interesting opportunities for research, by tracing how humans' capabilities interact and rely on one another. I examine one such interaction here: between children's social learning and open-ended play.

Learning facilitated by other individuals, known as 'social learning' (Kendal et al. 2018; citing Heyes 1994), has received attention from researchers for over a century (Woodward 1982). Humans' capacity for social learning has been described as 'the secret of our success' (Henrich 2016), and the importance of this transmission of information has led children to be described as 'cultural magnets' (Flynn 2008). Many researchers argue this human-unique capacity for social learning is neccessary for the cultural ‘ratchet’ effect (also called ‘cumulative cultural evolution’), which provides humans with a means of improving on cultural traits so that they become more beneficial for their users (Mesoudi & Thornton 2018). Meanwhile, recent research describes ‘play’ as a developmental context in which children explore novel solutions to future ecological problems (Pellegrini 2009). Play can be defined as a context in which children do follow some rules but also have a degree of freedom in both interpretation of the rules and choice of the activity’s constituents (e.g., its goals) (van Oers 2013). In educationalist literature, researchers tout play as the lynchpin for children’s learning and development (e.g., Golinkoff, Hirsh-Pasek & Singer 2006; also see Frost, Wortham & Reifel 2012). It is even argued that play was foundational for the evolution of human learning (Palagi, Stanyon & Demuru 2015). Nevertheless, the otherwise wide-ranging literature on social learning pays little attention to a crucial condition of play: its open-endedness (Rook 2008). Thus there is currently little data on whether the degree to which a social learning experiment is open-ended, versus close-ended, affects the copying behaviour of participants. Even Rook (2008), whose study makes progress in illuminating adults’ treatment of socially provided information in an open-ended, “creative” task, does not experimentally compare copying between open-ended and close-ended conditions, and does not explore the specific importance of play for children’s learning and development (e.g., see Bateson 2014). I posit that overlooking open-ended play may be masking variation in children’s flexible use of information derived from social and so-called ‘asocial’ learning. I therefore present my research in which children’s building with building blocks was used as a context to explore how children use information provided by other agents when the children are given relative freedom in the means *and* the ends of their activities.

In Chapter 5 I explore children’s use of socially provided information when the task given to them was close-ended (i.e., when children were given more detailed instruction). To identify variation in children’s use of socially provided information, I employed a distinction between ‘macrostructure’, the overall design of what is being built, and ‘microstructure’, the means by which ‘what is being built' is achieved by patterns in arranging building blocks. These two aspects of structural complexity are useful for making predictions about what information children may copy in some contexts rather than others. In Chapter 6 I explore how, in an open-ended task, children used information learnt socially in their building in some contexts but not others. To do this, I required stimuli which would elicit differential copying of macrostructure versus microstructure. I therefore turned to 'evidence of failure', the fact that children's building often involves instances of collapse. Previous research indicates that children recognise when copying from an unreliable 'model' (i.e., an unreliable individual) is non-beneficial, incentivising greater reliance on 'asocial' learning (Pinkham & Jaswal 2011; Birch, Vautheir & Bloom 2008; Bandura 1986; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017). However, previous studies also indicate that children can attempt to recreate what a failed model tried to achieve, or that children copy the model despite the evidence of failure (Meltzoff 1995; Want & Harris 2001, 2002; Sanefuji et al. 2004; Huang & Charman 2005; Carr, Kendal & Flynn 2015). Thus, the current thesis tries breaking a dichotomy between 'to copy' versus 'not to copy' down into contexts in which children may copy macrostructure alongside 'asocial' learning of microstructure, or may copy microstructure alongside 'asocial' learning of macrostructure.

The research presented in this thesis extends existing research in two ways. First, the thesis presents data from controlled experimental conditions which enable comparison of copying between open-ended and close-ended tasks. Second, the thesis explores how children's play behaviour could involve flexible combination of 'socially copied' alongside 'individually learnt' information, across two aspects of structural complexity. While data presented below do not give support for children's increased flexibility in copying microstructure design relative to macrostructure design in the open-ended task, data do give support for differences in children's copying behaviour between the close-ended and open-ended tasks. (1) Children in the open-ended task did not reliably copy social information more when the social model was successful rather than unsuccessful, which children in the close-ended task did. (2) Children in the open-ended task did not copy social information more when they encountered greater evidence of failure in their own task, as prior literature from close-ended tasks indicates they would. Given the sample size and experimental controls, these results are interesting and deserve further research in order to understand them.

Chapter 2: Theoretical context

Chapter 2 outlines previous work conducted on the two topics which my study aims to combine: (1) social learning and (2) children's play and creativity. In this chapter I argue that much contemporary social learning research relies on participants' performance in close-ended tasks, in which the goals of their activities are predetermined. I then argue that these close-ended tasks can lead to the impression that copying and creativity work competitively rather than collaboratively. Eight hypotheses, split between Chapters 5 and 6, are then introduced to test for differences between children's building between close- and open-ended tasks, and between children's copying of microstructure and macrostructure design.

2.1: Copying and the social learning literature

A fundamental question for this study is: what is 'copying'? I define it for the present thesis as the similarity of a learner's building outcomes to those of a model, when a learner is exposed to the activities of a model in some way. This recalls what is referred to as 'imitation' in older literature (e.g., Baer, Peterson & Sherman 1967), where it is used in a broader manner than the term is normally used today. This exposure of a learner may be to the activities of a model, or the effects of the model's activities (see Heyes 2001). Since a learner's building outcomes can be more or less similar to that of a model, it is conceptually plausible to say that copying can occur to greater and lesser extents. When a learner is exposed to the activities of a model, the extent of copying thus depends on how similar the learner's outcomes are to the model's. So while 'not to copy' refers merely to a relatively lesser degree of similarity of an observer's build to a model's, 'asocial learning conditions' here refers to the absence of a specific social cue, the absence of which makes it more difficult to copy the model. While the term 'asocial learning' is used here, since it is the standard term in the literature (Kendal et al. 2005; Kendal et al. 2018), these conditions are never actually 'asocial' in that they do provide social cues which can influence children's building. This means, however, that because a learner can have greater or lesser exposure to the model's activities, it is thus also conceptually plausible to say that the *possibility* to engage in copying can be present to greater and lesser extents. The extent to which copying is viable, even if it is desirable, is dependent on the information provided by the model (see Whiten & Ham 1992).

It is thus important to emphasise that the difference between 'to copy' and 'not to copy', as used in the present study, is not synonymous with the difference between 'to learn' and 'not to learn'. If a child "decides" that a social model does not provide them with useful information and "chooses" not to copy the model, the child's 'not to copy' behaviour may appear similar to how children in an 'asocial learning' condition will 'not copy'. However, learning induced by a model's evidence of failure (i.e., learning 'not to copy') is still a response to a social cue. Learning 'not to do' is not the same as 'not learning' (see Darby & Riopelle 1959). It is also important to note that the quotation marks used above indicate that the "choices" of a child may not be explicit, in that a child's "decision" to copy or not may be subconscious (Kendal et al. 2018). In sum, the choice between 'to copy' and 'not to copy' is a choice between two social learning options. Furthermore, the 'asocial learning condition' is only asocial insofar as it provides less (or vaguer) social information than the 'social learning condition'. Additionally, and as argued below, a learner's judgment of a model's usefulness is not necessarily uniform across microstructure and macrostructure (see Table 1).

Table 1
Decision table of children's copying from a model's building.

Decision	Microstructure design	Macrostructure design
To copy	Microstructure copying	Macrostructure copying
Not to copy	No microstructure copying	No macrostructure copying

Note. The decision whether to copy or not is split into two, for microstructure and macrostructure design. A child can plausibly combine either of the two options for each of the two columns.

Since copying is thus defined irrespective of the processes of social learning, it has no bearing on whether the similarity in building outcomes is achieved via specific social learning processes like 'imitation' or 'emulation' (Wood, Kendal & Flynn 2013b). In accordance with definitions of social learning processes used by Heyes (2001; Huang & Charman 2005), imitation is the copying of activities (i.e., motor patterns) and emulation is the copying of effects of activities. In the past, some researchers (e.g., Tomasello 1996; Call & Tomasello 1995) upheld a clear division between imitation (underpinned by a human-specific ability to understand the psychological causes of others' behaviours; see Tomasello 1990; Wellman, Cross & Watson 2001) and other, lower fidelity, social learning processes (also available to non-humans). However, many researchers argue that both imitation and emulation can be augmented by an understanding of others' intentions (e.g., Whiten & Ham 1992; Subiaul, Patterson & Barr 2016; Whiten et al. 2004), it is now accepted that chimpanzees are capable of understanding others' psychological states to the extent of a false-belief task (Krupenye et al. 2016), and there is some evidence that cumulative cultural elaboration can take place in the absence of imitation (Reindl et al. 2017; Caldwell & Millen 2009).

Since it is possible to imagine both the microstructure and macrostructure of a build being copied using either or both processes, whether information is copied via imitation or emulation is thus immaterial to the research question. Indeed, it has proven difficult to separate the roles of emulation and imitation in experimental conditions (Call & Carpenter 2002). Even under 'ghost' conditions, emulation can only be compared to the combination of imitation and emulation (Thompson & Russell 2004). Some researchers go further by rejecting such a distinction between imitation and emulation. Stout and Hecht (2017) argue that a dichotomy between information learnt imitatively and information learnt emulatively is not sufficient to document the complex differences between information learnt at different levels of goal organisation, or skills which entail various relations between ends and means depending on the specific context. Rather than a dichotomy, Stout (2011; citing Bekkering & Prinz 2002) instead describes a continuous hierarchy of goal-oriented actions, whereby children can copy (or not copy) a range of actions which are to varying degrees superordinate and subordinate to each other. Differences between imitation and emulation do not therefore already demarcate the microstructure-macrostructure division, and are not necessarily best placed for assessing the information children may or may not copy from a model's building.

Instead, the main distinction I make in children's copying of information is between 'macrostructure', associated with *what* is built, and 'microstructure', or the *way* it is built. This is a rather crude outline. There are clearly many ways to specify structural complexity, different intentions that go into building, and different information a learner might use from observing that building (Stout 2011). Nevertheless, the reason these two units of macrostructure and microstructure are adopted here is that they are already used implicitly in social learning literature. Moreau and Engeset (2016) categorise tasks as either more well-defined or ill-defined based on the degree of knowledge a participant has of (a) the aim of the activity, (b) the strategies available to achieve this aim, and (c) the initial state of the activity. Close-ended tasks distinguish microstructure from macrostructure by pre-determining this aim, or ultimate goal, of the learners' activities (i.e., resticting variation in macrostructure), thus leaving learners only with variability in the way to achieve this goal (i.e., the microstructure variants available to them).

Macrostructure copying is consequently associated with copying a model's 'ultimate' goal (whether or not a learner realises the model's ultimate goal with the same microstructure, similar to how 'goal emulation' is sometimes distinguished from 'imitation'; see Whiten & Ham 1992; Subiaul, Patterson & Barr 2016; Whiten et al. 2004). In the absence of a predetermined goal, selection of macrostructure design is an open-ended problem, as there is no predefined end-state which is correct, or even optimal. In the absence of parameters to define optimality, the child is faced with an axiological problem to invent such parameters themselves (Moreau & Engeset 2016). Microstructure can thus often be seen as a proximate goal. Once some sort of macrostructure is being worked towards, selection of microstructure is a close-ended task. This is because microstructure designs are more or less expedient for accomplishing a given macrostructure design. A child has a limited set of microstructure variants, that is, ways of putting blocks on top of each other. There are thus several different microstructure variants which a child could use to achieve a macrostructure which is similar to a model's. Therefore, a key point is that the social learning strategies applied to macrostructure do not specify the same learning strategies applied to microstructure. The macrostructure/microstructure distinction can thereby be used to explore beyond mere presence or absence of copying behaviour, towards an approach which emphasises degree and variation in participants' copying (see Barrett, Peterson & Frankenhuis 2016).

Many contemporary social learning studies use close-ended tasks. This means that they give participants a problem with a number of solutions, and study how participants are able to achieve the goal set for them (e.g., Flynn, Turner & Giraldeau 2016; Rendell et al. 2011; Wood et al. 2013a, 2013b). This means that much contemporary social learning literature explores participants' copying of proximate techniques to achieve predetermined ultimate goals. For example, Caldwell and Millen's (2010) spaghetti task determines a type of structure, a tower, and the aim, to build the tallest. By constraining macrostructure, this experiment investigates microstructural differences between participants' towers. Even tasks referred to as 'open-ended' in the literature (see Reindl & Tennie 2018) provide participants with a definite goal to achieve, against which they are judged more or less successful.

Ecological, or external, validity is the degree to which experimental results index conditions which exist beyond the experimental setup, in the 'real world' (Henrich, Heine & Norenzayan 2010). This lack of studies with open-ended setups creates a deficiency of ecological validity in the child social learning literature (Rook 2008), since the contexts in which children learn and explore their environment are often without an imposed direction to structure their activities (Bateson & Martin 2013; Gopnik 2012). However, some social learning studies do control the activities of learners less. For example, Meltzoff (1995; see also Bellagamba & Tomasello 1999; Carpenter, Call & Tomasello 2005 for similar studies with children of similar age) uses a more open-ended setup by allowing 18-month-old children to play with objects after adults performed evidence of failure (without ostensive cues such as vocalisation or facial expression), and by not rewarding them or reacting specifically positively if they succeeded in transforming the object. However, this experiment's open-endedness is still limited as the tasks they use do constrain the children's activities, in that the materials children are given to explore have specific goals built into them. Nevertheless, they found that children reproduced what the adults were attempting to do rather than simply copying their actions. This result indicates processes in children's learning in open-ended settings which are not explained by a simple distinction between copying a model versus not copying a model.

Research into 'creativity' has, at least until recently, demonstrated a comparable lack of attention given to the role copying takes in solving open-ended tasks, with only scarce consideration of the roles that copying behaviours could play in creative, open-ended task performance (Rook 2008). Rook (2008) explains that the reason for this is that most researchers in the field diametrically oppose copying with creativity (e.g., Sternberg 1999; Amabile 1996); i.e., copying and creativity compete and do not collaborate. I argue that the close-ended setup of many experiments can help to reinforce this dichotomy between 'to copy' versus 'not to copy' in the social learning literature too.

To understand the dichotomy, it is useful to look at Morin's (2016) description of two approaches to learning 'biases', which, in this context, describe the tendency of an individual to copy one thing but not another (see also Kendal et al. 2018). The 'strong' interpretation of social learning biases employs a strong distinction between 'to copy' versus 'not to copy'. This is because this strong perspective views biases as cognitive decision rules activated in the presence of formal domains. These formal domains are specific cues in a learner's environment, which operationalise latent social learning strategies to best fit the context (Laland 2004). Biases thus denote a heuristic system of hereditarily canalised social learning strategies (e.g., Henrich 2016; Richerson & Boyd 2005). An analogy would be a lock-and-key mechanism, with the correct key (i.e., environmental cue) opening the appropriate door (i.e., genetically specified social learning strategy). Cues in the learner's environment either activate one of the specified social learning strategies or they do not. Thus, in this strong interpretation, biases work on a strong dichotomy between 'to copy' and 'not to copy' (e.g., Whiten & Flynn 2010). This can arguably lead to oversimplified understandings of copying behaviour (Nurmsoo, Robinson & Butterfill 2010), which may prematurely oppose 'copying' to 'not copying'. Specifically, in social learning studies which focus solely on microstructure copying (i.e., studies which predetermine macrostructure design with a close-ended task) and participants do not appear to copy this microstructure, the researcher may then assume that copying in general is not exhibited in response to the given stimuli or context. Thus, a central concern of the present study is: are the cues that disincentivise copying at the microstructural level also capable of incentivising copying at the macrostructural level, and vice versa?

In contrast to the 'strong' interpretation, Morin (2016) describes the 'weak' interpretation of learning biases in which 'biases' merely denote the statistically common responses of human cognition to different types of ecological information. Biases, in this interpretation, do not entail claims as to their developmental origins or cognitive processes underpinning them (see Kendal et al. 2018). For example, biases would not have to adhere to a strict division between information from 'social' versus 'asocial' sources. While useful information may be provided directly by social models, such as demonstrating an effective way to build with blocks, this is not necessarily the case. Models may, equally, offer information in indirect ways. In Caldwell and Millen's (2009; see also Reindl et al. 2017) paper plane experiment, they examine children's social learning through their interaction with merely the effects of a model's activities – without any contact with the models themselves. In fact, any interaction with any environment is 'social' in as far as the environment is shaped by the activities of others (see Lewontin 1983). Sterelny (2012) calls this the 'downstream effects of niche construction'. In the present study, even in the absence of a model, the blocks children play with do not occur 'naturally'. Children, even in so-called 'asocial learning conditions', interact with a social environment constructed by the activities of conspecifics. Indeed the difference between more 'open-' versus 'close-ended' tasks is a prime example of the social framing of ecological conditions. A simple distinction between 'social' and 'asocial' information is thus complicated.

Correspondingly, a clear psychological difference between 'social' and 'asocial' learning also becomes fuzzy. Modelling work indicates that the adaptive value of the information gained from social learning is vulnerable to ecological change (Rogers 1988; Boyd & Richerson 1995). Truskanov and Prat (2018; see also Enquist, Eriksson & Ghirlanda 2007; Fogarty 2018) thus introduce 'trial and error' as a cognitive strategy to keep cultural traditions relevant to their contexts. The implication is that cognitive mechanisms for 'asocial' learning be considered intrinsic to the evolutionary success of social learning. Indeed Morin (2016) does not assume that the hypothesised cognitive mechanisms for the faithful transmission of cultural information are different from those for individual reinvention (e.g., Heyes 2017, 2012; Perreault, Moya & Boyd 2012; citing Heyes 1994; Plotkin 1988). In this perspective the difference between 'to copy' and 'not to copy' in any given context is weak, in that the difference between the two is not necessarily cognitively profound. Instead, the difference between copying and not copying may merely be a matter of the direction of the same cognitive mechanisms toward different targets. This is not to say that a weaker interpretation of social learning biases espouses a non-modular mind, but that modules (however defined) are organised around types of activity rather than a strict division between 'social' versus 'asocial' cues (Morin 2016; see Kalish, Griffiths & Lewandowsky 2007). There is currently no conclusive evidence for the extent to which the cognitive mechanisms for 'social' and 'asocial' learning are different or the same (perhaps excluding those for language and teaching; Kendal et al. 2018). In any case, the weaker version of the social learning bias framework is useful as a critique of the assumptions of some social learning research, and particularly this research's reliance upon the premise of a clear opposition between copying and not copying.

2.2: Play and creativity

I have argued that issues of ecological validity, caused by the lack of open-ended children's social learning studies, mean that social learning research might overlook important features of children's copying behaviour. This is because play behaviour is widely acknowledged to be an important scaffold in the ontogeny of various animals (Bateson & Martin 2013), especially humans (Gopnik 2012), and especially in contexts of tool use and traditional (i.e., informal) education (Lancy 2017). Furthermore, contemporary definitions of play specifically emphasise a relative lack of externally imposed constraints on activity: they emphasise open-endedness. van Oers (2013) employs activity theory to describe human play as a recognisable form of activity. They typify play through highly involved agents, who (implicitly or explicitly) follow some rules, but who have a degree of freedom in (a) interpretation of the rules and (b) choice of the activity's constituents (e.g., its goals). Open-endedness is thus not a lack of rules or goals for activity, but a context in which such rules and goals can be more readily re-interpreted by the activity's participants. Definitions of human play from Garvey (1990), Gray (2013), Pellegrini (2009), Stuart Brown (2010), and Weisberg et al. (2013) each also emphasise play as a child-led activity or one which is not prescribed by goals external to the activity (Zosh et al. 2018). Play is thereby described as a social niche, or institution, in that it is a context defined by (a relative lack of) external constraints and incentives on activity (van Oers 2013; Leont'ev 2009; Vygotsky 1967; see Sinha 2015, citing Goodwin & Goodwin 2004; Yamagishi 2011).

While people have, at times, considered play a diversion with little benefit for individuals' development, this is no longer the case among researchers (Athey 1984; see Golinkoff, Hirsh-Pasek & Singer 2006 for a discussion of the history of ideas about play). Bateson (2014) describes the prevailing opinion of play as a context for fine-tuning of motor skills and neuromuscular systems, and outlines play contexts as safe spaces for experimentation in behavioural responses to environmental stimuli (see also Gopnik 2012; Cook, Goodman & Schultz 2011). Bateson (2014) further explains play as a probing device to explore beyond 'locally optimal' solutions to ecological problems. If this experimentation took place outside of play, children could suffer if their novel solutions were non-beneficial. If, in play, they did not suffer such consequences, then eventually behavioural solutions with greater optimality than the local optima could be found and deployed in non-play contexts (Pellegrini, Dupuis & Smith 2007; Pellegrini 2009). Thereby, play may be seen as a social-developmental niche in which affordances for activity are manipulated to help scaffold childhood ontogeny (Zosh et al. 2018; see Vygotsky 1978; Leont'ev 1978; Flynn et al. 2013; Whiten & van de Waal 2018). An experimental setup that is close-ended is thus not reflective of childhood play (see Bateson & Martin 2013) in which children may copy not only microstructural solutions to problems, but also the macrostructural aims which produce those problems. For example, a child may copy either the overall structure of a tower, or the way building blocks are put together in building the tower, or they may copy both the tower *and* the structure by which the tower is built. Therefore, it is necessary to study explicitly how manipulating the open- or close-endedness of social learning tasks affects children's copying of microstructure and macrostructure.

Research into creativity has recently made some progress in exploring the gap between copying and innovation (e.g., Rook & van Knippenberg 2011; Mecca & Mumford 2013; Moreau & Engeset 2016; Bonawitz et al. 2011; Rook 2008). The most crucial feature of 'creative' tasks is their open-endedness (Amabile 1996), or a relatively lesser degree of clarity with which participants know the structure and aims of their activity (Moreau & Engeset 2016). This relative degree of open-endedness corresponds well to the recent recognition that children's play constitutes a spectrum from the absence of any kind of goal, as in 'free play', to more structured and goal-oriented play, as found with children's games (Zosh et al. 2018). In these creative tasks, 'creativity' is usually considered to be the production of innovative (i.e., in some way original) and elegant solutions (Mecca & Mumford 2013; citing Besemer & O'Quin 1999; Christiaans 2002). Copying, meanwhile, is often dealt with through social comparison theory (Festinger 1954; Bandura 1986), in which models act as comparison standards which influence agents' activities (Rook 2008). It would thus seem apparent that the greater the degree to which a participant does not copy, the more original, and thus creative, the product of their activities will be (Marsh, Landau & Hicks 1996). This is the opposition between copying and creativity Rook (2008) describes. However, again paralleling the social learning literature, research actually provides a mixed picture of how exemplar copying relates to creative problem solving (Mecca & Mumford 2013). Some studies do indicate that copying inhibits creativity (e.g., Smith, Ward & Schumacher 1993; Smith & Blankenship 1991; Weisberg 1986). Yet others indicate that copying of exemplar solutions contributes to creativity (e.g., Rich & Weisberg 2004; Weightman 2007). This would further indicate that individuals may not merely choose to copy or to not copy given a context, but that there are interesting questions to be explored in how copying behaviour is deployed within given contexts (see Marsh, Landau & Hicks 1996).

Specifically, the 'creativity' of an artefact cannot be defined merely by the degree to which it is different from the artefacts that came before it. If the artefact is not fit for purpose, or is much more 'inelegant' than what came before, then it cannot be called a truly 'creative' solution (Biro, Haslam & Rutz 2013). Mecca and Mumford (2013) address this by positing that creative solutions could be found through modification of existing solutions, producing solutions which are both informed by the prior work of others, and yet extend their work by manipulating it in certain ways. They thus focus on (and find support for) strategies by which adults modify information presented to them, and the patterning of these strategies according to the type of information presented. This ties in with research by Carr, Kendal and Flynn (2016; Reader & Laland 2003) on 'innovation by modification'. The study of individuals' use of social information in open-ended creative tasks thus appears to be a developing area of research. The current study therefore brings research on the topic of creativity into a closer synthesis with child social learning studies so that the role of social information in open-ended tasks can be better understood.

In the primate literature, juveniles are thought responsible for the majority of innovations, and technical innovations in particular (Perry, Barrett & Godoy 2017; Whiten & van de Waal 2018; though see Reader & Laland 2001). However, it is generally held within the human creativity literature that children under the age of around 8 years are not competent innovators of creative solutions to problems (Carr 2016), a finding with wide support from studies of 'WEIRD' and 'non-WEIRD' children (Sheridan et al. 2016; e.g., Cutting, Apperly & Beck 2011; Chappell et al. 2013; Nielsen et al. 2014). For example, roughly half of 8-year-olds tested can fashion a hook to retrieve a basket, and very few children of 5 years of age are able to (Carr 2016). It is a surprising finding given that children of the same age range are demonstrably able to selectively copy the very same solutions from models, compared to less effective solutions. Children of this age range, even as young as 4 years old, can understand the mechanics that mean a bent pipe cleaner is a more suitable tool than a straight one, since when they are presented with both they regularly choose the hook (Beck et al. 2011). While a couple of studies (Subiaul et al. 2015; Tennie et al. 2014) do arguably present some evidence for young children's innovation, these studies employ less stringent definitions of 'novel behaviour' (Carr 2016).

In an attempt to solve this dilemma, Sheridan et al. (2016) find evidence that children's creativity is influenced by contextual factors including the experimental setting and setup. They suggest that in environments in which children are given greater freedom to experiment with materials, children of 4 to 7 years showed higher levels of tool innovation than reported in previous studies. These results suggest that contextual factors are important influences on children's creative problem solving. If creativity is linked with social learning, as many researchers suggest (e.g., Legare & Nielsen 2015; Morin 2016; Heyes 2012), then this degree of task open-endedness may well be an important factor in children's copying behaviour. Indeed, Riede et al. (2018; citing Palagi, Stanyon & Demuru 2015) argue that children's play provides space for creativity and innovation which contrasts with other research findings of a high degree of conformity and conservatism when children are exposed to social information. However, to my knowledge the extent to which children's reliance on social information in open-ended tasks resembles or differs from that in close-ended tasks has not been established experimentally. The experimental setup, described below in Chapter 3, therefore aims to test this by comparing children's copying between close-ended and open-ended tasks.

2.3: Encounters with evidence of failure

Acquisition of information about the physical world is not the only purpose social learning behaviours are thought to serve (others include, for example social affiliation and communication; Clay, Over & Tennie 2018; Nielsen & Slaughter 2007). However this 'instrumental' purpose is widely considered important (see Legare & Nielsen 2015; Tennie et al. 2009; Užgiris 1981). One facet of the child social learning literature thus explores how children learn about the world through interaction with an environment which does not align with either their intentions or those of an individual they observe (e.g., Horner & Whiten 2007; Want & Harris 2001). 'Intention' is here taken to mean the hypothesised cognitive causes of a given bodily action (see Searle 1983). These unintended dissonances between intention and realisation are here termed 'evidence of failure'.

If two children are playing with building blocks, there is a good chance that at some point one of their constructions will accidentally collapse. This collapse demonstrates to the child that some aspect of their building has failed, due to the dissonance between whatever they were trying to do and the fallen blocks surrounding them. This is 'internal' evidence of failure; the dissonance is between the perception of the material environment and the child's intentions. When their own, non-copied solutions are shown to be ineffective in close-ended experiments, children have an increased propensity to copy from others (Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a). In close-ended tasks, when risk of failure is artificially increased relative to a control baseline, children engage in greater copying (Caldwell & Millen 2010). This makes evolutionary sense, since learning directly from conspecifics only produces benefits when it can increase an individual's ability to accomplish something in a given task relative to asocial learning (Boyd & Richerson 1985). So if non-copied information is shown to fail, *relative* risks of relying on copied information decrease (Feldman, Aoki & Kumm 1996). Thus, prior literature using close-ended tasks expects that children exposed to internal evidence of failure should copy more.

'External' evidence of failure represents a similar dissonance, but this time between the learner's perception of the material environment, and the learner's inferences of a model's intentions. By at least five years of age, children recognise when copying a model is not expedient (Want & Harris 2001; Bijovet-van den Berg 2013; Rakoczy, Tomasello & Striano 2004). Children rely on more non-copied information when a model's information is shown to be inaccurate or inefficient, or the model has an 'unreliable' reputation (Clement, Koenig & Harris 2004; Ma & Ganea 2010; Pinkham & Jaswal 2011; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017; Birch, Vautheir & Bloom 2008; Bandura 1986). This also accords with evolutionary theory, which expects copying to be beneficial to individuals, and sustainable at a population-level, only when a learner can select to copy information which is useful (Giraldeau, Valone & Templeton 2002; Kendal et al. 2005; Truskanov and Prat 2018; Enquist, Eriksson & Ghirlanda 2007; Whitehead & Richerson 2009). There is clear evidence that children are, at least in some contexts, "optimal-" rather than "over-" imitators (Evans et al. 2017).

However, children still use information provided when an experimental model's performance displays evidence of failure (Meltzoff 1995; Want & Harris 2001, 2002; Sanefuji et al. 2004; Huang & Charman 2005; Carr, Kendal & Flynn 2015). Moreover, existence of the phenomenon of 'over-imitation' itself suggests that children (and adults) can be surprisingly oblivious to the risks of copying. 'Over-imitation' is defined as the copying of information which is redundant to the instrumental goal (Whiten et al. 2016). Explanations of the phenomenon often invoke the social role of copying behaviour introduced above (see Clay, Over & Tennie 2018; Over & Carpenter 2012). A lack of causal understanding of the task (on the part of the learner) also seems to have an influence, suggesting that over-imitation serves to transmit information which is useful but not understood (Lyons et al. 2011; see Burdett et al. 2018 for evidence of interaction between both 'social' and 'causal understanding' explanations). There is also a long history of findings from other animals that social learning can actually be enhanced through exposure to unsuccessful models (Templeton 1998; citing Herbert & Harsh 1944; John et al. 1968; Beauchamp & Kacelnik 1991). It therefore appears that *whether* children copy or do not copy under a given condition can be formulated as a more nuanced question: *what* information is it that children copy and what information is it that they do not copy? The experiment described below in Chapter 3 therefore employs the distinction between microstructure and macrostructure to test for children's copying of different kinds of information under experimentally manipulated conditions.

The child, an active participant (Flynn et al. 2013) in the social-developmental niche of play, may combine different information from copying and 'asocial' learning in flexible ways. Presented with a model's external evidence of failure, a child may be disincentivised to copy some aspects of the model's building but also incentivised to use other information. Imagine two children playing with blocks; one begins to build a tower, but the tower falls down. The other child may then try to build a similar tower because there are no disadvantageous consequences of failure, and therefore manipulation of environmental affordances in this way becomes a playful challenge. Since the first child demonstrates external evidence of failure in the way they go about the task, the second child may try to achieve the tower using a different microstructure design (similar to what some social learning literature refers to as 'goal emulation'; Whiten & Ham 1992; Subiaul, Patterson & Barr 2016). Macrostructure may thus be copied independently of microstructure. This flexibility could also be observable in children's playful responses to internal evidence of failure. Bateson and Martin (2013) argue that play is about finding more optimal solutions to ecological problems. After the child's tower collapsed, they may therefore not give up on their goal. This is because evading the problem would not find a solution to it. Instead, they may copy different information from their friend, such as a more stable way to put the blocks together (i.e., copy their microstructure design) so as to achieve their original goal by different means. In this case, microstructure may be copied independently of macrostructure.

2.4: Rationale for hypotheses

To introduce the hypotheses, it is useful to have an understanding of the eight conditions under which data was collected from participants. The participants were aged between 5 and 11 years old. Table 2 describes these conditions: the model was either 'social' or 'asocial', and either 'successful' or 'unsuccessful', whilst the task was either 'close-ended' or 'open-ended'. While an 'asocial model' may seem anachronous, it is 'asocial' in line with the definitions introduced above. While the social model provided information of use to a participant, the asocial model did not. Builds produced in the 'asocial' condition could then be compared with those built in the presence of a 'social' model, enabling me to judge the degree of similarity between participant and model builds that arose in the definite absence of copying. Furthermore, in addition to these conditions, each participant demonstrated lower or higher 'internal evidence of failure' in their own building.

The purpose of Chapter 5 is to explore how microstructure and macrostructure copying is affected by a close ended task, as used in previous social learning studies. This aim is addressed by comparing the similarity of children's buildings to the social model across various sources of variation. The first hypothesis concerned how macrostructure copying varied between the close-ended and open-ended tasks. In the close-ended task, children were given a specific aim to achieve: they were told to build the tallest tower they could. In the open-ended

Table 2

The eight conditions under which children's building with wooden blocks was studied.

Condition	Model demonstration	External evidence of failure	Task type
1	Building blocks ('Social')	Successful model	Close-ended
2	Building blocks ('Social')	Successful model	Open-ended
3	Building blocks ('Social')	Unsuccessful model	Close-ended
4	Building blocks ('Social')	Unsuccessful model	Open-ended
5	Irrelevant toy ('Asocial')	Successful model	Close-ended
6	Irrelevant toy ('Asocial')	Successful model	Open-ended
7	Irrelevant toy ('Asocial')	Unsuccessful model	Close-ended
8	Irrelevant toy ('Asocial')	Unsuccessful model	Open-ended

task, however, children were not given a specific goal to achieve: children were told to build the 'best building' they could.

Hypothesis 1 I predicted that children, given the close-ended task, would build macrostructure designs which were more similar to that of the social model than when they were given the more open-ended task. This effect should be observed irrespective of whether or not children had access to the social model (which demonstrated this macrostructure design). See Carr (2016) and Legare et al. (2015) for prior examples of the effectiveness of such verbal instruction.

Two more hypotheses were made about how microstructure copying is affected when macrostructure is constrained by the close-ended task. This replicates studies discussed above which use close-ended experiments. Hypothesis 2 investigates whether children built differently when a 'social' model was present (i.e., a model who built with building blocks) rather than an 'asocial' model (i.e., a model who performed an irrelevant activity). Hypothesis 3 investigated whether children copied this social model more when the model built successfully compared to when the social model's building frequently collapsed

Hypothesis 2 When given a close-ended task, I predicted that children who had access to a relevant 'social' model would build microstructure designs which were more similar to that of the social model than children who observed an irrelevant 'asocial' model. I expected this to hold true also when the social model was unsuccessful in building. This would be compatible with the previous finding that people undertaking creative tasks conform to exemplar types even unconsciously (Smith, Ward & Schumacher 1993), called 'design fixation' (Rook 2008; Shalley & Perry-Smith 2001).

Hypothesis 3 When given a close-ended task, I predicted that children who had access to a successful social model would build microstructure designs more similar to the social model than children who observed an unsuccessful social model. This is because the successful, but not unsuccessful, model showed the microstructure design to be useful in tower building (Pinkham & Jaswal 2011; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017).

The fourth hypothesis of Chapter 5 investigated the effect of the close-ended task on how similar participants' macrostructure designs were to the social model's macrostructure design.

Hypothesis 4 When children were given a close-ended task and a social model, I predicted that children would not show increased macrostructure similarity to the model when the social model was successful rather than unsuccessful. This is because it was expected that the close-ended task would constrain macrostructure diversity.

The purpose of Chapter 6 is then to investigate how children balance copying of microstructure and macrostructure within open-ended task – a more play-like context. I aimed to create conditions in which children combined social learning of microstructure with so-called 'asocial' learning of macrostructure, and conditions in which children combined social learning of macrostructure with 'asocial' learning of microstructure. The first two hypotheses used the success of the social model to predict variation in the similarity of children's builds to the social model's build.

Hypothesis 1 I predicted that when children built in an open-ended task and observed a social model (i.e., a model which was relevant to their task), children would show greater microstructure similarity to the social model when this model was successful rather than unsuccessful. This would be the same effect as found in prior studies using close-ended tasks, which indicate that children react to unreliable or inaccurate models by relying less on social information (Pinkham & Jaswal 2011; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017).

Hypothesis 2 I predicted that children building in an open-ended task with a social model would not show increased macrostructure similarity to the social model when the model was successful rather than unsuccessful. This is because children may infer the unsuccessful model's macrostructure intentions and try to emulate them (Meltzoff 1995; Huang & Charman 2005; Carr, Kendal & Flynn 2015). External evidence of failure was expected to be associated with no weaker macrostructure similarity to the social model because the open-ended task approximated playful conditions. Since research indicates that children use play to experiment with solutions to ecological problems (Bateson 2014; van Oers 2013; Pellegrini 2009), the failure of the model may well have incentivised the children to explore different solutions to achieve the goal that the model did not.

The final two hypotheses of Chapter 6 then used participants' internal evidence of failure to predict variation in the similarity of children's builds to the social model's build.

Hypothesis 3 I predicted that in an open-ended task, children who encountered more failure in their own building (i.e., greater 'internal evidence of failure') would copy the microstructure of the social model more than those with less internal failure. This would be the same effect observed in prior close-ended experiments, in which encounters with failure in their task cause children to defer to a social model (Williamson, Meltzoff

& Markman 2008; Wood, Kendal & Flynn 2013a; Caldwell & Millen 2010).

Hypothesis 4 I predicted that internal evidence of failure should have had no effect on degree of macrosructure similarity to the model in the open-ended task. In open-ended conditions, in which children 'choose' not only the means of a given activity but the ends (van Oers 2013), children may appropriate a social model's microstructure for their own macrostructure designs. This is because open-ended play is about experimentation (Bateson & Martin 2013; Gopnik 2012). When internal evidence of failure is encountered, to copy a model's macrostructure would be merely to avoid the problem rather than solve it. Children exhibiting higher internal evidence of failure may therefore 'choose' to build macrostructure designs which are different from the social model.

Furthermore, there are three additional factors which may have influenced each child's microstructure and macrostructure similarity scores: age, sex, and their degree of attention to the experimental video. First, younger rather than older children were predicted to show higher microstructure and macrostructure similarity scores with a social (rather than asocial) model, since literature indicates that they are more limited in their capacity for innovation than older children (Carr 2016; Beck et al. 2011; Chappell et al. 2013), that younger children are more faithful copiers than older children (Carr 2016), and that younger children display greater social conformity compared to older children (Walker and Andrade 1996).

Second, Brand, Brown and Cross (2018) find that when asocial learning was risky, females but not males were more likely to rely on social information. Furthermore, there is some evidence from non-human animals pointing to greater reliance on social learning by females: Vale et al. (2017b) find female chimpanzees adopt socially provided task solutions more readily than males, Reader and Laland (2001) find males apes more likely to engage in innovative activities than females (though see Hopper et al. 2014 on chimpanzees), while tests with rodents suggest females engage in greater social learning than males (Ervin et al. 2015). Also, sex differences between what children build may also be apparent due to different expectations of what girls and boys should play with and achieve (Freeman 2007; Brahms & Crowley 2016; see also Ehrlinger & Dunning 2003). I therefore adopted the tentative hypothesis that, with a social model, males would demonstrate lower microstructure and macrostructure similarity scores than females.

Thirdly, there is evidence that stimuli which are more “attention grabbing” are copied more than other stimuli (Berger 2011; Mesoudi, Whiten & Dunbar 2006; also see Whitehouse 2004; Davenport & Beck 2001). I therefore predicted that children with higher scores of ‘attendance to the experimental video’ would show higher microstructure and macrostructure similarity scores when the model was social (rather than asocial).

Chapter 3: Methods of data collection and analysis

Chapter 3 presents an overview of and justification for the methods used to collect experimental data in a non-laboratory setting. It then moves on to discuss procedures of data coding and the Bayesian methods used to analyse these data.

3.1: Data collection

Data collection was conducted in the Centre for Life, a science centre in Newcastle upon Tyne (UK). This allowed me to make use of Durham Anthropology Department's relationship of collaborative research with the nearby Centre (see Kendal et al. 2016). Over school holidays and weekends from 23 October 2017 until 18 February 2018, I recruited 659 participants from family visits, of which data were usable for 565 samples. Ninety-four participants were excluded after data collection either because they were too old or too young for the primary school age range (i.e., younger than 5 or older than 11 years old), or because they observed the building of other children taking part in the experiment. The non-laboratory experimental setup breaks down distinctions between controlled experiments and 'real-world' observations (see Rudman et al. 2017; Sheridan et al. 2016). It thus attempts to maximise and balance both internal and external (i.e., ecological) validity. I could therefore experimentally manipulate several variables, and assess my hypotheses by comparing children's behaviour across these manipulations. Yet from the perspective of the child participants, their participation represents interaction with just one science centre exhibit out of many other non-data collecting exhibits. This provides my study with an insight into children's playful interaction with such a science centre exhibit, an example of children's 'real world' playful learning (Kendal et al. 2016). Non-laboratory settings, such as science centres and museums, thus offer opportunities to balance experimental control with 'real-world' behaviour (see Rudman et al. 2017; Sheridan et al. 2016). As there were entrance fees for the Centre for Life, £8 for adults and £6 for children aged between 5 and 17, there may have been certain demographic biases, as well as the geographic bias of situating the experiment in northeast England with its WEIRD participants (Nielsen et al. 2017; Henrich, Heine & Norenzayan 2010). Nonetheless, by taking the experiment to the participants, this study was able to reach people who would be less likely to volunteer their children for laboratory-based experiments. Written consent was provided for children by the parent/guardian to confirm that they understood and agreed to the experiment (see Appendix 1 for an example of the information and consent forms). The child could withdraw from the task at any time. The Durham Anthropology Department Research Ethics Committee granted permission for the experiment to go ahead on 26 June 2017, with final ethics permission granted on 25 January 2018.

The research reported in this thesis relied on the eight experimental conditions displayed in Table 2 (page 31). All conditions involved children who were given building blocks to play with. The children, tested individually, were told they would watch a video five minutes long, during which time they could build with the blocks. The children knelt at a low table, which enabled children to build tall structures without them being limited by how high they could reach (see Figure 1 for a photograph of the experimental setup).

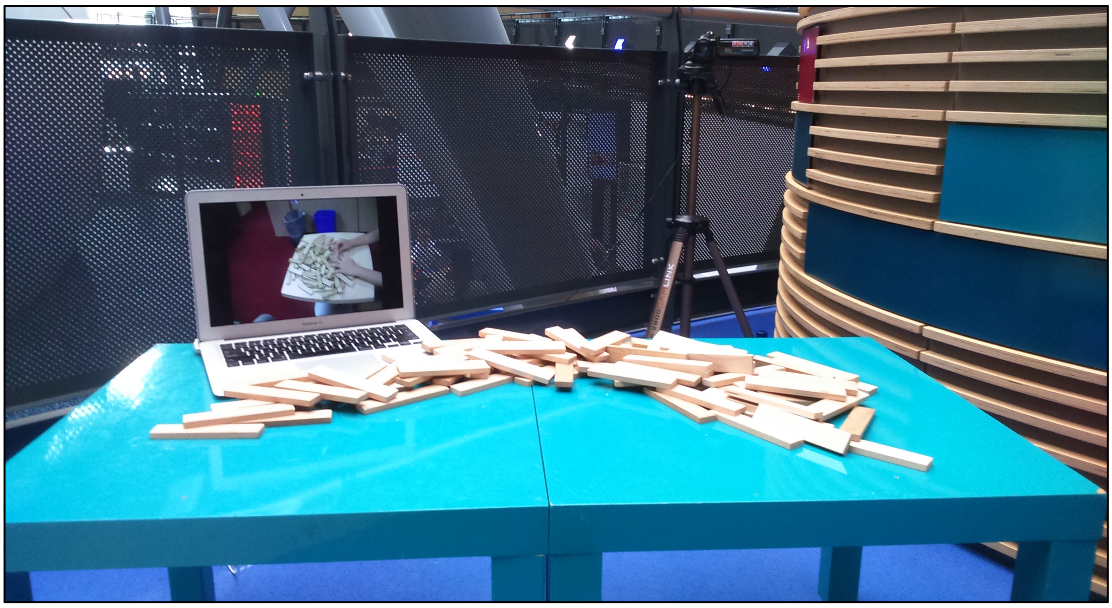

***Figure 1*. Photograph of the experimental setup. Children built on the blue tables using the wooden blocks. They observed the experimental video on the laptop on the left of the photograph, and their activities were recorded by video camera on the right. Another photograph, from a slightly different angle, can be found in Appendix 2.**

One variable which differentiated the conditions was the extent to which children were verbally constrained in their building. 276 children (49% of participants) were given a 'close-ended' task in which they were told to build the tallest tower that they could. This was intended to restrict their use of the blocks' affordances in macrostructure building. Children given the close-ended task were told:

> "I'm going to show you this video. The video lasts about five minutes. In those five minutes, I want you to build the *tallest* tower you can. You can build it any way you like, but the aim is to build something as *high* as possible [gesturing with hands]. The only thing is: you do have to watch the video, because you can *only* build for as long as the video is playing. So when the video stops, it means you have to stop building, and so you have to keep an eye on the video so that you know when to finish. Does that make sense?"

The italics indicate particular stress placed on the words. Where children could not speak English (fewer than ten children in total), their guardians translated the information and the children's answers. This approximates other social learning setups which give learners a specific goal to accomplish with specific resources (e.g., Caldwell & Millen 2009). Such verbal instructions, followed by visual observation, have previously been found able to influence children's copying behaviour (Carr 2016; Legare et al. 2015; see further citations in Clay, Over & Tennie 2018).

The 'open-ended' task, meanwhile, did not verbally constrain the child to building a given macrostructure. Instead, 289 children (51% of participants) were told:

> "I'm going to show you this video. The video lasts about five minutes. In those five minutes, I want you to build the *best* building you can; you can build *anything* you want; you have *total* freedom. The only thing is: you do have to watch the video, because you can *only* build for as long as the video is playing. So when the video stops, it means you have to stop building, and so you have to keep an eye on the video so that you know when to finish. Does that make sense?"

Therefore the open-ended task still provided a problem for the children; they had to build the 'best building'. But they had to decide not only how to achieve this goal, but also what the goal – the 'best building' – would be. This was a relatively more open-ended condition than the 'close-ended' task, in that children had a greater role in interpreting the aims of their activities. The variety of things which the children said they built indicates that they did not only take 'building' or 'tower' to mean a copy of a real building, though the wording conceivably had some influence since various kinds of towers and buildings accounted for 68% of the things children said they built (386 out of 563 responses). However, there was no control group to compare this with what children would say they built when given instructions which do not mention terms like 'tower' or 'building'.

These two tasks thus represent two points on a continuum from more close-ended to more open-ended (see Moreau & Engeset 2016; Zosh et al. 2018). Indeed it can be considered that the mere presence of a model, even including the so-called 'asocial' model, reduces the open-endedness of an experiment by influencing the participants' activities. Moreover, it is important these conditions did approximate a 'playful' environment in order to encourage the children to behave playfully (Bateson 2014). While the actual extent of open-endedness did experimentally vary, other parameters of the playful environment were held constant. Children in the close-ended condition were still presented with a relatively playful context in that, beyond internal dissatisfaction of failing to achieve the specific goal, there were no external costs to failure. This created a more conservative test of my predictions.

So that no specific type of play was rewarded and incentivised, there was no prize which unsuccessful children were deprived of, as there is in some experiments (e.g., Caldwell & Millen 2009). In addition, children may well have associated the building blocks themselves with playful contexts. A questionnaire was administered to guardians (N=68) of children at the Centre for Life between February and March 2017. Fifty four percent of the children (N=37) had reportedly played with physical blocks in the past week, 16% (N=11) in the past month, and 15% (N=10) in the past year. The children aged between 4 and 13 years old (mean=7.7, standard deviation=2.45), with 38% female and 62% male. These data indicate building blocks were a common toy for children for this age range visiting the Centre for Life. The structure of the building blocks allowed the children to use them in a large variety of ways, offering a range of affordances for building which they could choose between (see Gibson 1986). The building blocks were also causally transparent – they contained no hidden ‘traps’ to surprise the children, as sometimes seen in social learning experiments (e.g., Dean et al. 2012). This is significant since a lack of causal transparency is thought to contribute to ‘over-imitation’ (Lyons et al. 2011).

Another variable that distinguished the experimental conditions was the nature of the model. Of 565 children, 149 participants (26%) observed a model who built a tall tower without any evidence of failure. The model completed the tower at the five-minute mark, after building steadily and consistently throughout the video. 148 participating children (26%) observed the same model attempt to build a tower but encounter multiple instances of severe evidence of failure. Throughout the five minutes the model built the tower up, it collapses, they try again and fail repeatedly. Both of these conditions were therefore ‘social’ model conditions, the sample size of which adds up to 297 participants (53% of the usable dataset). Finally, 268 children (47% of the usable dataset) observed the same model bounce a table tennis ball on a table tennis paddle. These were the ‘asocial model’ conditions, since the children could derive no information of use to tower building from them. The children thus had to build a tower without immediate social cues about how to go about it. The reason this ‘asocial’ condition also contained a social agent is to account for the possibility that children’s building could be affected merely by the presence of a video displaying another social agent. Furthermore, to account for the possibility that children’s building could be influenced by the success of the model even in ‘asocial’ conditions, participants were either placed into an ‘asocial successful’ (124 children, 22%) or ‘asocial unsuccessful’ (144 children, 25%) condition, in which the model either bounced the ball without evidence of failure or with evidence of failure, respectively.

The model's identity and dress, and the method of filming were held constant between the different conditions, and the videos for each condition filmed on the same day. To obscure the view of the model's body and face, the videos were all shot from an angle roughly 90° from the model's face of direction and the videos focused only on the model's hands and bare forearms. This lack of information about the model's face and body language also minimised the risk that children took the model's activities to be pedagogical (see Csibra & Gergely 2011; Heyes 2016). The true identity of the model was a 25-year-old woman. Out of 559 respondents, 53% (N=294) said the model was a girl, 39% (N=220) said the model was a boy, and 8% (N=45) said they did not know. This trend holds in seven out of eight conditions, with only participants in the Unsuccessful Open condition saying the model was a boy (53%, N=39) more frequently than a girl (35%, N=26). Out of 559 respondents, 47% (N=262) thought the model was 15 years old or younger, while 42% (N=232) thought the model was over 16. This leaves 12% of respondents (N=65) saying that they did not know the model's age. There were three conditions for which the difference between the percentage of participants saying the model was over 15 and the percentage saying the model was not over 15 was greater than 15%: participants in both Unsuccessful Close with 63% (N=44) and Unsuccessful Open with 58% (N=42) saying the model was 15 years old or younger, and Asocial Successful Close with 45% (N=31) saying the model was older than 15 years old. Given the spread of sex and age guesses, I believe it is fair to assume that participants did not have a clear idea of the model's sex or age, with a slight tendency to guess the model was younger than the reality.

Children were encouraged to observe the video but were not specifically encouraged to either copy or not copy the model. This was accomplished by telling the children that they had to "keep an eye on the video so you know when to finish [building]." Of 565 participants, 63% (N=358) played for at least 295 seconds (4 minutes 55 seconds). The mean length of time for which children played was 268 seconds, with a standard deviation (henceforth 'SD') of 63 seconds. There was little variation between conditions: from Unsuccessful Open with the greatest mean time, 286 seconds (SD=45 seconds), to Asocial Unsuccessful Close with the lowest mean time, of 255 seconds (SD=73 seconds). The lack of variation in the time children spent building is important, as it indicates that testing time and exposure to the model were relatively constant between the various social and asocial conditions (see Reindl & Tennie 2018 for the importance of this). Although children were also told that the video would last about five minutes, data presented below indicates that this technique did encourage children to attend to the video. This attendance to the video stimulus can be measured by counting the frequency of glances children made at the laptop playing the video (which was positioned on the tabletop which children were building on). For this purpose, the participants were themselves videotaped whilst building (see Figure 1 and Appendix 2 for photographs of the positions of the laptop and video camera). The frequency of eye gaze direction towards the stimulus video was not intended to measure directly children's degree of social learning; for instance, children could monitor the video in their peripheral vision. Instead this measure merely suggests the degree of interest which children demonstrated towards each video stimulus. Limitations of the eye gaze direction data are that it did not record the duration of children's eye gaze direction towards the video, and occasionally children's faces became obscured from the camera. Gaze duration was not measured due to limitations in time, and because this measure is likely to result in greater error in coding since it was sometimes unclear for exactly how long a child looked towards the laptop.

However, as a rough measure of children's attention to the video stimuli, Table 3 displays the relatively invariable degree of attendance shown to the video model between the different conditions (both social and asocial) across mean scores, standard deviations, and ranges. The video to which the highest degree of attention was given was 'Unsuccessful Close', with a mean of 7.69 glances given per minute, while 'Unsuccessful Open' was the video with the lowest attention score, at 5.91 glances per minute. The standard deviations, however, of these two means overlapped substantially, at 4.21 and 3.65 glances per minute respectively.

Table 3
Participant attendance to the video across the eight conditions of the experiment.

Condition	Mean frequency of glances to the video per minute	Standard deviation of the mean	Lowest score	Highest score
Successful Close	6.97	4	1.37	18.72
Successful Open	7.12	4.68	1.25	25.51
Unsuccessful Close	7.69	4.21	1.4	21.61
Unsuccessful Open	5.91	3.65	1.13	15.2
Asocial Successful Close	7.59	4.9	1.38	25.51
Asocial Successful Open	6.95	4.79	0.98	23.7
Asocial Unsuccessful Close	6.55	4.24	0.42	16.59
Asocial Unsuccessful Open	7.27	5.61	0.79	23.33

This demonstrates the success of the time limit strategy to encourage attention to the video stimulus, since it was able to encourage children to pay equal attention to useful and non-useful videos. Differences between the similarity of children's builds to the model builds under the different conditions are thus unlikely to be due to some videos being more eye-catching or engaging than others, though the possibility for this was included in the modelling strategy outlined below.

The measure of 'internal evidence of failure' was collected at the time of children's building, by tallying the frequency of collapse in each child's building across three categories of severity. I thus measured both the prevalence of collapse and the extent to which each collapse was catastropic. It was thus assumed that both a greater prevalence of collapse and a greater severity of collapse contributed to children's judgement of the degree to which they were doing well in the task. Category 1 severity was defined as a collapse in which less than a third of the total number of blocks were moved from the positions they were put by the participant. As the total number of blocks available to the child was 100, a category 1 collapse affects less than 33 blocks. Accordingly, category 2 severity was defined as a collapse affecting between 34 and 66 blocks, and category 3 severity affecting more than 66 out of the 100 blocks. Since not all children were given the stated goal to build the tallest tower, I could not use the height of their builds to measure the degree of internal evidence of failure. The collapse-based measure of internal evidence of failure is thus useful since, assuming that children did not build with the intention to create collapse, it is universal across various building styles and goals.

To create a singular measure of the degree of internal evidence of failure, the frequency of collapses were summed together in a method in which a category 1 collapse confered one point, category 2 collapse confered two points, and category 3 collapse confered three points. This created an ordinal range of positive integers that reflected both the frequency and severity of collapses across all of the participants. This number was then divided by the length of time for which each participant built for, to account for the fact that one participant may have a greater internal evidence of failure score merely for building for a longer time. Inter-coder analyses were performed by a postgraduate student familiar with quantitative methods on a sample of 113 participants' video recordings (re-coding 20% of the 565 original internal evidence of failure samples in the usable dataset). To assess the difference between the two coders' internal evidence of failure scores I used a weighted quadratic Cohen's Kappa, a variant of the Kappa designed for ordinal data and in which a greater distance between the two coders' ratings is penalised more than a lesser distance (Cohen 1968). This returned a statistic of κ=0.71 (standard error=0.07), with the 98% confidence interval ranging from 0.56 to 0.86. Thus the greatest probability for the agreement between the two raters fell into the 'substantial' agreement category (from 0.61 to 0.8) of Cohen's original (1960) framework (see McHugh 2012). This provides confidence in this variable, as McHugh (2012) recommends that data for medical clinical studies can be accepted with a Cohen's Kappa above 0.60. They (McHugh 2012) also recommend reporting the standard percentage agreement between coders, as the Kappa can excessively lower the agreement estimate. Taking only exact matches between the two raters' scores, 61.1% of ratings were in agreement. This rises to 77% agreement if a difference of two steps is tolerated, and 85% of each coder's ratings were within 3 ordinal steps of each other.

I counterbalanced for sex across the conditions. Participants' sexes were identified by their guardians, who were asked to state whether the child was 'Male', 'Female', 'Other', or otherwise there was an option to decline to state the child's sex (see Appendix 1 for a copy of the questions). One child's guardian declined to state the child's sex, which I coded as 'N/A'; no guardian identified their child's sex as 'Other'. The total number of male participants was 305, accounting for 54% of the usable samples, while females numbered 259, accounting for 46% of the usable dataset.

There was, however, some variation among experimental conditions in the numbers of males and females. This is displayed in Table 4. The condition with the fewest number of females relative to males was with the successful social model and close-ended task ('Successful Close' in Table 4), in which females numbered 25 out of 71 participants (i.e., 35% of participants). The condition with the highest relative number of females, meanwhile, was with the asocial successful model and open-ended task ('Asocial Successful Open' in Table 4), in which females numbered 36 out of 64 participants (56% of participants). Nevertheless, such variation can be accounted for if it is included in statistical analysis. Even the conditions with the greatest imbalance across sexes have adequate sample sizes of both males and females. This permits any differences

Table 4
Variation in the number of children identified as 'female' between the eight experimental conditions.

Condition	Percentage of female participants	Number of female participants	Total number of participants
Successful Close	35.21%	25	71
Successful Open	39.74%	31	78
Unsuccessful Close	54.05%	40	74
Unsuccessful Open	44.59%	33	74
Asocial Successful Close	48.33%	29	60
Asocial Successful Open	56.25%	36	64
Asocial Unsuccessful Close	43.66%	31	71
Asocial Unsuccessful Open	46.58%	34	73

in behaviour between sexes to be assessed statistically. A variable was therefore included in statistical analyses which represented ‘male’ with a ‘0’ and ‘female’ with a ‘1’. This enabled any effects of being female, rather than male, on the outcome variable to be accounted for.

I also ensured a range of participant ages across the UK primary school age range (i.e., 5- to 11-year-olds). I chose to study children of primary school age because they are old enough to understand and manipulate causal relationships in their physical surroundings. Horner and Whiten (2007) note that by the age of 3 years children can understand the physical principles of contact (Bates, Carlson-Lunden & Bretherton 1980), force (von Hofsten et al. 1998), and gravity (Hood 1995). 3-year-olds are also able to combine these principles to predict the outcome of causal events (Bullock, Gelman & Baillargeon 1982). However, Horner and Whiten (2007) also note that, according to their experiment, it appears that while 3- and 4-year-olds have knowledge of these complex causal relationships they find it difficult to apply them consistently in practice. Instead, Horner and Whiten say that it is by the age of 5 or 6 years that children could successfully navigate their experiment’s causal relations consistently. Since the experiment I use here was dependent on children’s understanding of physical causality and effects of principles such as friction and gravity, I required participants with the cognitive abilities to deal with the challenges that arise in building. By the age of 5 years, children also have a well-developed ability to reason about others’ intentions. Carpenter, Akhtar and Tomasello (1998) found that infants even between 14 and 18 months old were able to discriminatively imitate a model’s intentional actions in preference to their accidental actions, signposted by the model’s verbal exclamations. By five years of age, children are able to understand both the ‘prior intentions’ and ‘intentions in action’ of a model (Astington 1991; see Searle 1983; Astington & Gopnik 1991; all cited in Meltzoff 1995). Moreover, by the age of five years, children’s proclivity to copy is widely reported (Clay, Over & Tennie 2018; e.g., Clay & Tennie 2017). After each child completed the building task, I asked them to say how old they were (in years). Guardians also gave each child’s date of birth, which could be used to corroborate the children’s answers if required. The mean age of participants in the usable dataset was 7.8 years, with a standard deviation of 1.79, and a range between 5 and 11 years old. The condition with the lowest mean age, Unsuccessful Close, at 7.5 years old (SD=1.83), has a standard deviation which overlaps the condition with the highest mean age, Asocial Unsuccessful Close, at 8.4 years old (SD=1.85). Furthermore, the similar standard deviations for these mean ages indicate a similar spread of ages between the conditions.

A photo was taken of each child's build after the five minutes for building had elapsed, or when the child themselves wanted to stop building. When a photo could not be taken, for example the building collapsed at the end of the five minutes or in the moments after it, a still image was taken from the video recording. This means that some of the photos used to compare the similarity of children's buildings represent a stage of the building process prior to the end of the five minutes. This practice was adopted since an unfinished building is a more useful piece of data than a mess of fallen blocks, which, by my definition of evidence of failure, does not reflect what the child was trying to do. If I had excluded those buildings that collapsed, I would have biased the sample towards children who built more stable structures. These photos were then used to code the degree of similarity each child's build had to either the successful or unsuccessful social model builds, across the two levels of microstructure and macrostructure.

3.2: Data coding

I now move on to discuss the methods employed for data coding, starting with the microstructure similarity score. A microstructure score of relative similarity to the model was assigned to each child's build. A microstructure similarity score of 0 represents no microstructure variants in common between a child's build and the model's build, and a score of 16 represents all microstructure variants in common between a child's build and the model's build. The microstructure similarity score for a given child's build was created by comparing the microstructure variants it shares with a model build, from a list of 16 variant types (see Appendix 3 for the list of microstructure variants, with some further explanation). For example, a participant's build could incorporate blocks positioned on their long/broad side, their long/thin side, or their short/thin side. Furthermore, the participant's build could incorporate blocks overlapping each other at a roughly 90° angle, a roughly 180° angle, or a roughly diagonal angle. The microstructure variants employed by the two social model builds were the same regardless of condition; the differences between the 'successful' and 'unsuccessful' social models were only in the frequency of evidence of failure, not in the microstructure variants employed. Simply, the microstructure similarity of the successful social model build to the unsuccessful social model build was 16 – there were no microstructure variants one had which the other did not. This same set of model microstructure variants could therefore be applied for microstructure similarity scoring of children's builds in the 'asocial' conditions. The microstructure similarity scores of participants' builds in the asocial conditions thus measure the similarity of children's builds to a model they did not see: they measure the similarity that arises between builds in the absence of the possibility of direct copying. The microstructure 'similarity' scores of builds in the 'asocial' model conditions, therefore, are a crucial baseline since microstructure similarity can plausibly occur two ways: (1) children directly copying what they see in a model, (2) indirect convergence towards the same variants through interaction with similar environmental conditions. Such environmental conditions may include the affordances of the blocks (Gibson 1986), prior experience with building blocks (see Carr 2016), verbal instruction to achieve a pre-specified goal (Carr 2016; Legare et al. 2015), as well as other more indirect social cues (see Lewontin 1983; Sterelny 2012; Morin 2016). Therefore, if significant differences were observed in the degree of microstructure similarity between the 'social' conditions compared with the 'asocial' conditions, it would suggest that these greater similarities would be due (at least in part) to copying behaviour.

This microstructure coding procedure represents an improvement on similar procedures used by previous literature. Rook (2008) codes structures which adult participants built from Lego blocks. One shortcoming of the coding procedure they employ is that they list a number of "random" features of the model structure, and then look for how many of these features are present in each of the participants' builds, without recording the number of features participants included that were not features of the model. This means that, in their study, a participant's build which only uses the features demonstrated by the model is only coded as similar to the model as another participant's build which combines all of the model features with several different ones of their own. My study improves this by counting the total number of differences, in a list of microstructure variants, between each participant's build and the model's build. While representing a step forward, the microstructure coding technique used here has its own shortcomings. One is that it coded microstructure variants as discrete types each equally different from each other. This may not always be the case, however, as some microstructure variants could be considered more similar than other variants. Also, this coding procedure did not take into account the frequency by which a build uses a specific microstructure variant. For example, a participant's build which mostly used blocks on their broad edges except for one block on its long thin edge, was as similar to a model which only used blocks on their long thin edges as another participant's build which used blocks on their long thin edges for all except one block on its broad edge. Such a within-build frequency measure of microstructure variants was not possible due to limitations imposed by time, though this sort of frequency measure could be applied in future studies.

A student with experience in quantitative methods conducted inter-observer reliability analysis for 19.82% of the microstructure similarity score data (112 samples), using the same techniques as described above for the internal evidence of failure variable. The microstructure similarity scores showed even less variation across the two coders than the scores for internal evidence of failure. The Cohen's Kappa statistic was κ=0.88 (standard error=0.02), with a 98% confidence interval ranging from 0.82 to 0.94. This indicates 'strong' agreement (McHugh 2012). In terms of simple percentage agreement, 85.7% of the 112 cases had exactly the same value between the two coders, 97.3% of the cases had values separated by at most a difference of two steps, and there were no cases with values separated by more than three steps.

Each participant's build was also given a score of macrostructure similarity to the model's build. The overall shape of the things which children built whilst observing a successful model (social and asocial) were compared with the image of the successful social model's build, and all of the things children built whilst observing the unsuccessful model (social and asocial) were compared with the image of the unsuccessful social model's build. This allowed the degree of relative macrostructure similarity of participant builds to model builds to be compared between the conditions. Since overall similarity is a relatively subjective measurement (simply, individuals' assessments of the macrostructure of a given build are likely to vary even when the build itself is the same), macrostructure scores must take account of multiple individuals' assessments. The final macrostructure similarity score was therefore an average of multiple coders' similarity scores. I used the website Prolific Academic (Prolific Academic Ltd. 2016) to recruit adult participants to code macrostructure similarity scores for each build online, an online research platform with advantages over alternative websites (Peer et al. 2017) and laboratory-based research (Woods et al. 2015). The Durham Anthropology Research Ethics Committee granted permission for this online experiment to go ahead on 25 January 2018. Participants rated the similarity of pairs of child and model builds on a Likert scale from 1 (more different) to 7 (more similar). In order to limit bias in the coding, the sequence of images to be compared was randomised for each participant, and the side of the screen occupied by model versus child build was varied. Due to a technical issue, two builds were not coded for macrostructure: one from Asocial Successful Open, and one from Asocial Unsuccessful Open. Participants could only complete the online experiment after giving explicit consent that they agreed to take part, and that they understood the data would be kept confidential, information identifying individuals would not be stored in the data, and that they would be free to withdraw at any time for any reason (see Appendix 4 for information and consent forms). Of 128 macrostructure coders, 50 (39%) were female and 78 (61%) male. Each participant coded 130 child builds, resulting in a mean sample size for each build of 30.3 (SD=5.1). For each build, the macrostructure similarity score was found by taking the modal Likert score from the range of ratings. There were 30 cases where a build had more than one modal score, and in most of these cases the two modal scores were consecutive. In these 30 cases, the mode closest to the mean Likert score for that case was chosen for analysis. The rationale for this was that where more than mode was present, the data did not clearly dictate which value to use. Therefore, analysis should make use of the datum which reflected best the spread of the other scores in each case, in preference to the datum which was further from the spread of the other scores.

3.3: Data analysis

Both macrostructure and microstructure scores were therefore ordinal outcome variables. Liddell and Kruschke (2018; see also Burkner and Vuorre 2018) outline the problems of analysing ordinal variables with metric models, i.e. models that assume data are on an interval or ratio scale. Problems include both false positive and false negative errors, as well as the inversion of the ordering of means, leading to systematic inversions of effects. I therefore used a strategy of multilevel ordinal categorical Bayesian analysis. The various benefits of a Bayesian analysis are described by McElreath (2016). The process of Bayesian analysis consists of updating the relative probabilities of ways in which data could have been created by parameters, as new data are presented to the model (McElreath 2016). An ordinal categorical model pairs a categorical likelihood function with a cumulative link function, allowing estimation of the effects of predictor variables across multiple levels of the outcome variable (McElreath 2016). Multilevel modelling thus allowed me to make the effects of variables dependent on the effects of other variables (McElreath 2016). Modelling was conducted in R Version 3.4.1 (R Core Team 2015) and Stan (Carpenter et al. 2017) via the RStudio interface (see Racine 2012), using the function 'map2stan' in McElreath's (2016) 'Rethinking' package. For all parameters of the models presented below, I used a generic weakly informative prior: Normal(0,1). I used this prior so as to rule out unreasonable parameter values, whilst allowing variation in parameter values which are more plausible (McElreath 2016). For all of the models, I checked Rhat and n_eff values as indicators of successful model convergence, measuring the degree of chain convergence and the chains' effective sample size respectively (McElreath 2016). All models used for graphing of data had n_eff values of at least 800 for each parameter (McElreath 2016 states that good estimates of posterior distribution can be taken with an n_eff value as low as 200). All models used for graphing also showed Rhat values for each parameter of 1, indicating full convergence (McElreath 2016).

Each of the models presented below was the product of a process of model comparison in which various combinations of predictor variables were tested to ascertain which combination provides the most useful way of understanding the data. Adding variables can add information which helps inform predictions about the data. Models with too little information to make useful predictions are said to be 'underfit' to the data. However, models can also 'overfit' the data (McElreath 2016). An overfit model is one that is so well fit to the present data that it becomes less useful for predicting new data. McElreath (2016) outlines Akaike weight and the difference in WAIC ('Widely Applicable Information Criterion') scores as a useful way to discriminate between different models which contain different combinations of parameters. Weight can be understood as an estimate of the probability that a model will make the best predictions on new data, relative to the other models being compared (McElreath 2016; Burnham & Anderson 2002). A smaller WAIC score also indicates better out of sample deviance, or the better a model is at predicting similar data it has not been fit to (McElreath 2016). The models below therefore represent the combination of predictor variables which produce the highest Akaike weight and lowest WAIC values.

For each hypothesis of Chapters 5 and 6, I report the Bayesian model's estimates for the posterior distribution of the hypothesised predictor variable. These figures, while useful in deciphering the positive or negative direction of a predictor's effect on the outcome, are problematic in inferring anything more. This is because, in an ordered categorical model, they describe changes on the level of log-odds across cumulative probability distributions (McElreath 2016). This is in turn because the model includes not just one 'intercept' parameter, but several – one for each of the thresholds separating the ordinal categories of the outcome variable. When microstructure similarity was the outcome variable there were 16 ordinal categories, in which each step away from '1' and closer to '16' represents increasing microstructure similarity of a participant's build to the social model's build. When macrostructure similarity was the outcome variable there were 7 ordinal categories: the product of the seven-grade Likert scale (a low number denoting low similarity, a high number denoting high similarity) which was administered to online participants. The effects of the predictor variables, and interactions between the predictor variables, on the outcome variable must therefore instead be visualised across all of the ordinal categories of the microstructure or macrostructure similarity scales. This was accomplished using variations on the triptych plot. These plots are predicted effects in that they are what the model expects new data to show if more data were gathered from the same experimental process. These predictions allow inference of trends in past data because the model creates these predictions by iteratively interacting with the data collected from the real experiment, through the Bayesian updating process.

As an example of interpreting the cumulative probability graphs, take Figure 2a. The blue lines represent 100 samples from a given model's posterior distribution. They illustrate the estimated effects of changing the predictor variable (represented on the horizontal axis) from a value of '0' (on the left of the horizontal axis) to a value of '1' (on the right of the horizontal axis). The blue lines thus lose height as they go from a value of '0' (left) to '1' (right). To understand what this means for the outcome variable (i.e., the similarity of a child's build to the model's build), we first must understand what the vertical axis describes. So look now only at the left-hand side of Figure 2a, where the blue lines meet the vertical axis. At the bottom of the vertical axis is '0', with '1' at the top. These values refer to cumulative probability, which starts from zero and sums to one at the top.

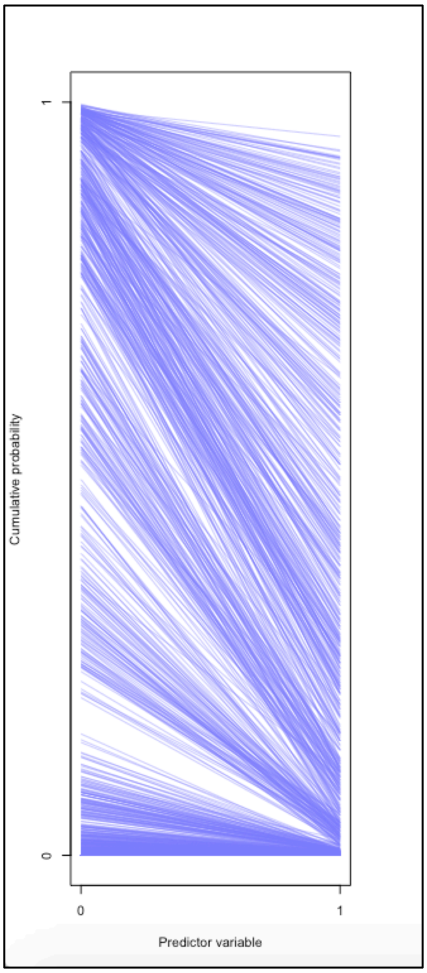

Figure 2a. **Example graph A. It illustrates the predicted effects of turning a predictor variable from '0' (on the left of the horizontal axis) to '1' (on the right of the horizontal axis). This effect is registered on the distribution of cumulative probability (on the vertical axis, from '0' to '1') across the ordinal categories of the outcome variable. This effect is represented by 100 blue lines, each of which is a sample from the posterior distribution of a Bayesian model.**

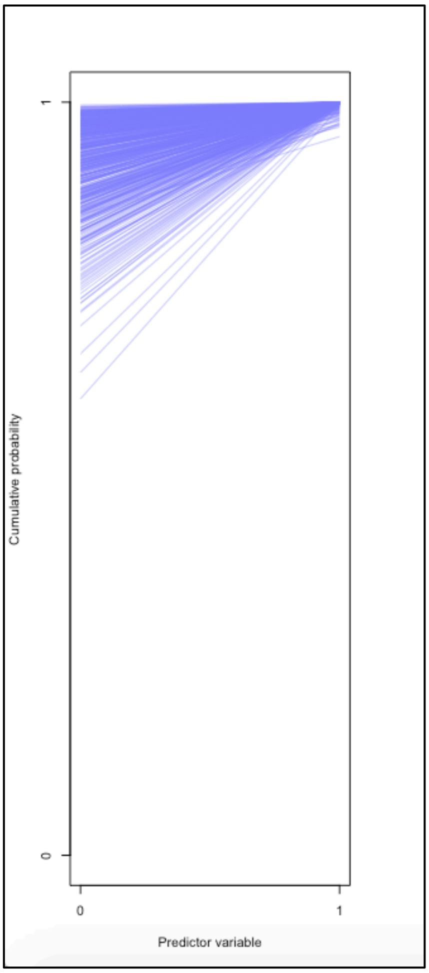

Figure 2b. **Example graph B. Graph B also shows the predicted effects of turning a predictor variable from '0' (on the left of the horizontal axis) to '1' (on the right of the horizontal axis) on the distribution of cumulative probability (vertical axis, '0' to '1') across the ordinal categories of the outcome variable (represented by 100 blue lines, each a sample from the posterior distribution of a Bayesian model).**

The vertical axis thus represents how the cumulative probability is spread across the total number of ordinal categories of the outcome variable. When the outcome variable is microstructure similarity, there are 16 categories which the cumulative probability is spread across. When the outcome variable is macrostructure similarity, there are 7 categories over which the cumulative probability is distributed. The aim of this analysis is to examine sources of variation between how much cumulative probability is allocated to the different categories: whether greater cumulative probability was taken up by the lower ordinal categories (i.e., builds were rated as less similar to the model) or whether greater cumulative probability was taken up by the higher ordinal categories (i.e., builds were rated more similar to the model). This job is done by the blue lines, which illustrate the Bayesian model's predictions for what would happen if the predictor variable on the horizontal axis was turned from 0 to 1. The 100 blue lines are thus, more accurately, 100 predictions from the Bayesian model of where the thresholds between the ordinal categories lie in terms of cumulative probability.

Both Figures 2a and 2b thereby represent the effect of a change in the predictor variable (horizontal axis) on the distribution of cumulative probability across the ordinal categories of the outcome variable (vertical axis). In Figure 2a, the blue lines descend as the predictor variable is changed from '0' to '1'. In an ordinary correlation graph, this would indicate a negative correlation. However, in this kind of graph, the descent of the lines represents a positive correlation. This is because the blue lines bunched at the bottom of the vertical axis mean that the outcome variable's lower ordinal categories take up *relatively less cumulative probability*. Figure 2a thus describes a change in the distribution of cumulative probability from 'predictor variable=0', where more of the cumulative probability is taken up by the lower ordinal categories, to 'predictor variable=1' where more of the cumulative probability is taken up by the higher categories of the ordinal scale. In other words, the Bayesian model predicts that the change in the predictor variable causes cumulative probability to shift up towards the higher categories of the ordinal scale, with participants' builds rated more similar to the social model's build when the predictor variable has the value '1'. Contrast this with Figure 2b, which differs from Figure 2a in two obvious ways. First, the blue lines in Figure 2b intersect with the vertical axis on the left of the graph much further towards the top (further towards '1'). This indicates that most of the cumulative probability is taken up by the lowest of the ordinal categories: when the predictor variable=0, participants' builds are mostly given low scores of similarity to the social model. Second, the blue lines ascend as they plot the change from 'predictor variable=0' to 'predictor variable=1'. In a standard correlation graph, this would indicate a positive correlation between variables. However, since the vertical axis measures the cumulative probability of the outcome variable's ordinal categories, rather than values of the outcome variable itself, this is not the case. The blue lines' ascent instead implies that by changing 'predictor variable=0' into 'predictor variable=1', participants' builds are rated as even less similar to the social model's build. With this introduction to data interpretation, I will now move on to the first of the results from data analysis.

Chapter 4: Investigating effects of age, sex, and attendance to social information

Chapter 4 presents the results of Bayesian data analysis for the effects of participant age (section 4.1), participant sex (section 4.2), and participant attendance to the experimental video (section 4.3). I will therefore start my analysis by assessing the degree to which these variables, and interactions between them, can predict variation in participants' microstructure and macrostructure similarity scores across all conditions of the experiment. Since these variables are not the main focus of my research, I will provide less of the details of data analysis than I will for Chapters 5 and 6. However, more detail, as well as descriptions of the models used, can be found in Appendix 6. Summaries of the model comparison processes for these models can be found in Appendix 5. The data for the microstructure analyses comprised 561 cases, while data for the macrostructure analyses comprised 559 cases. Four cases were dropped because they had missing values of 'age' and 'attendance to the video' in them, which causes problems for the statistical software (McElreath 2016), with two further cases dropped for macrostrucure analyses due to the issue in the macrostructure coding procedure described in Chapter 3.

4.1: Participant age

Across both microstructure and macrostructure variation, older children were shown to have created structures more similar to the social model than younger children. Copying appears to have played an important role in this difference between older and younger children since the positive effect of age on microstructure and macrostructure similarity scores was visibly strengthened by the presence of the social, rather than asocial model. Data thus challenge findings of prior literature that younger children display greater conformity to social stimuli than older children (e.g., Carr 2016; Walker & Andrade 1996).

The predicted real effect of participant age, in interaction with three other variables, on microstructure similarity scores are shown in Figures 3 and 4. The influence of the social model strengthened the positive effect of increased age on microstructure similarity scores. In all of the conditions with a social model (graphs C and D in Figures 3 and 5), the difference between younger and older children's microstructure similarity scores was greater than with the asocial model (graphs A and B in Figures 3 and 4). Across both low and high internal evidence of failure (Figures 3 and 4), the change from an asocial to a social model intensifies the already positive relationship between age and microstructure similarity in the asocial close-ended condition (i.e., between graphs A and C). Across both low and high internal evidence of failure (Figures 3 and 4), the change from the asocial to the social model also reverses the negative relationship between age and microstructure similarity found in the asocial open-ended condition (between graphs B and D). This is the opposite effect to what was predicted by prior literature (e.g., Carr 2016). More detailed description of the results of this statistical analysis, and the structure of the statistical model which generated it (Model 1), can be found in Appendix 6.1.

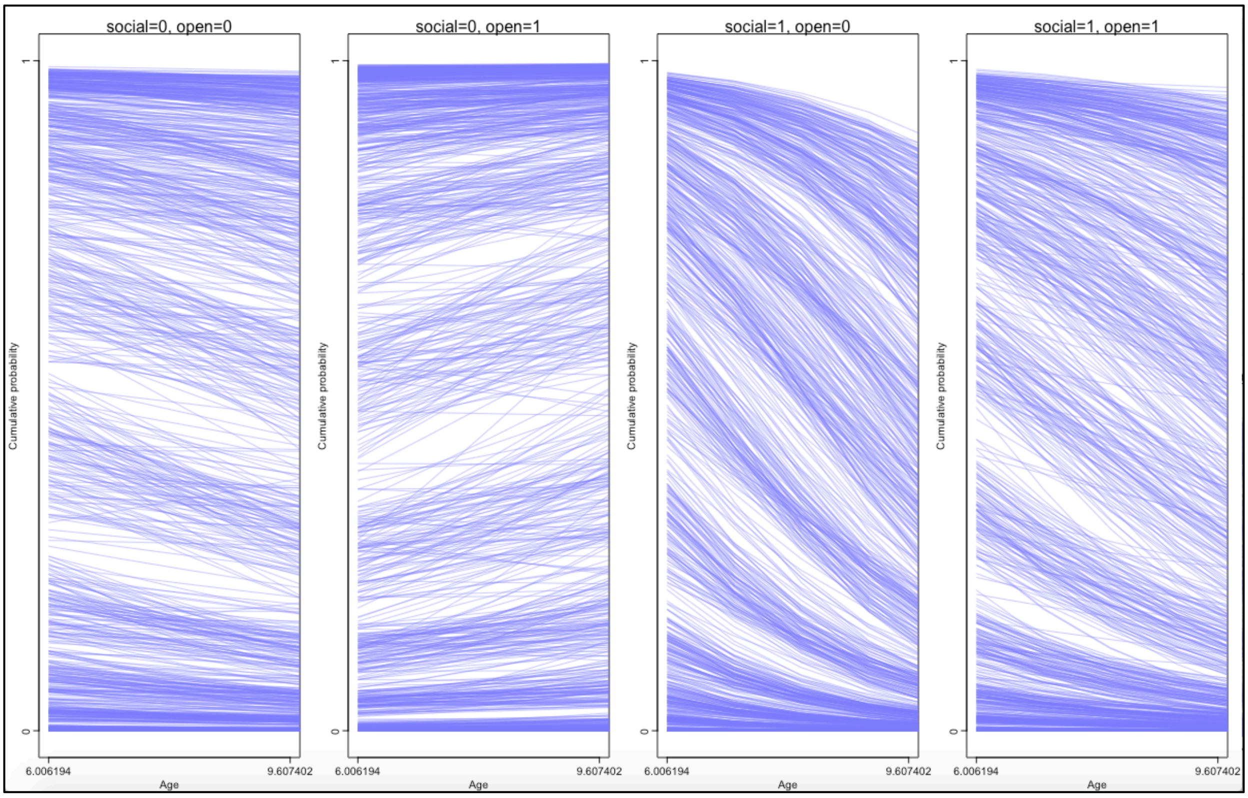

Figure 3. Four graphs (A, B, C, and D from left to right) illustrating Model 1's predicted effects of increased participant age on microstructure similarity scores. The 'low' age, on the left of each graph's horizontal axis, was set one standard deviation below the mean age of the entire usable dataset (mean=7.81), while the 'high' age was set one standard deviation above this mean and can be seen on the right of each graph's horizontal axis. Graphs A and B, on the left, show the effect of increasing participant age when the model was asocial, while graphs C and D, on the right, show the effect of increasing participant age when the model was social. Graphs A and C (far left and second from right) show the effect of increasing participant age when the task was close-ended, while graphs B and D (on the far right and second from left) show the effect of increasing participant age when the task was open-ended. For all four of these graphs, the degree of internal evidence of failure of the participant was set at 0.69: one standard deviation below the mean score for the entire usable dataset (4.62).

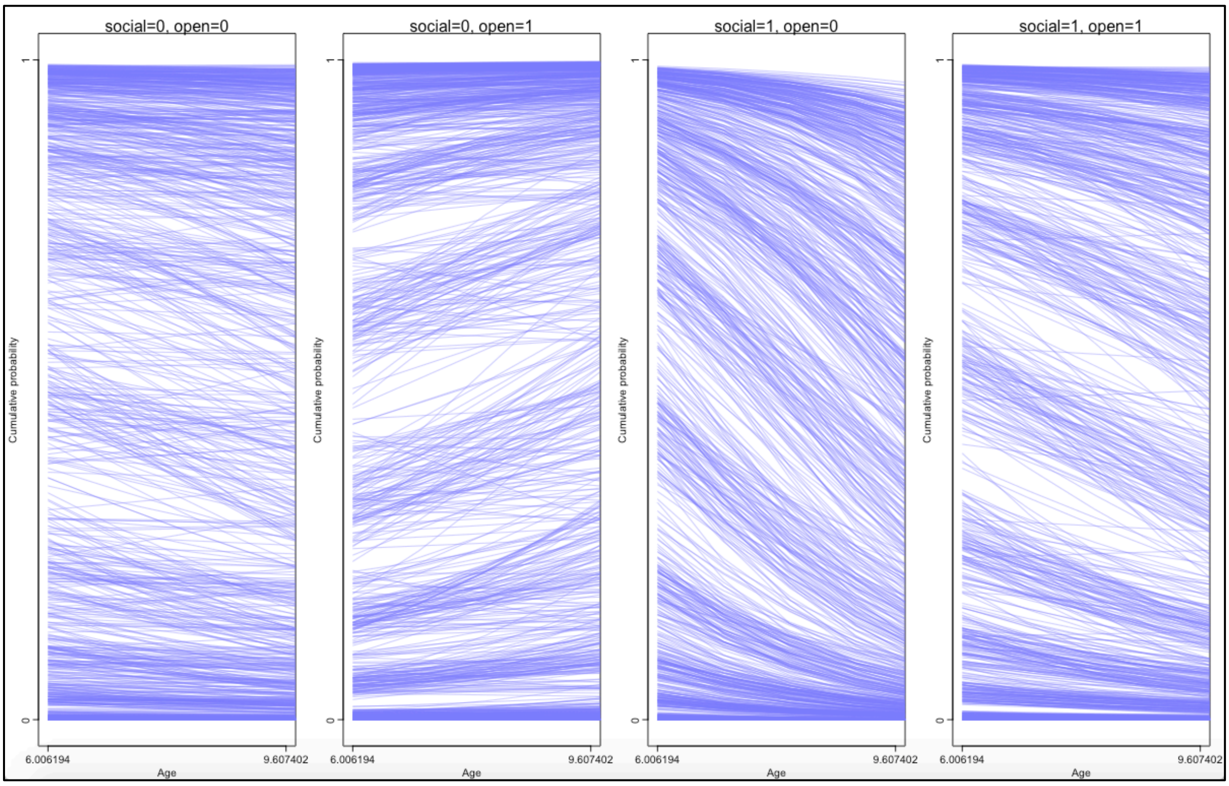

***Figure 4*. Four graphs (A, B, C, and D from left to right) illustrating Model 1's predicted effects of increased participant age on microstructure similarity scores. The 'low' age was set one standard deviation below the mean age of the entire usable dataset (which was 7.81 years old), while the 'high' age was set one standard deviation above this mean. The interactions described in the four graphs, between the asocial (graphs A and B) and social models (graphs C and D), and close- (graphs A and C) and open-ended tasks (graphs B and D) were the same as the graphs in Figure 3. However, for all of these graphs, the degree of internal evidence of failure of the participant was instead set at 8.54: one standard deviation above the mean score for the entire usable dataset (4.62).**

Figures 5 and 6 show the effect of participant age on macrostructure similarity scores to have been more uniform than that of the effect of participant age on microstructure scores (in Figures 3 and 4). In all of the eight interactions graphed there was a positive relationship between age and macrostructure similarity. The influence of the social rather than asocial model appears to have again strengthened the positive effect of age on macrostructure similarity. In the asocial conditions (graphs A and B of Figures 5 and 6), there was a visibly positive effect of age on macrostructure similarity, but this was significantly increased across the four social model conditions (graphs C and D of Figures 5 and 6). This again contradicts the stated hypothesis that younger children should show increased macrostructure similarity scores with a social model. More detailed description of the results of this statistical analysis, and the structure of the statistical model which generated it (Model 2), can be found in Appendix 6.2.

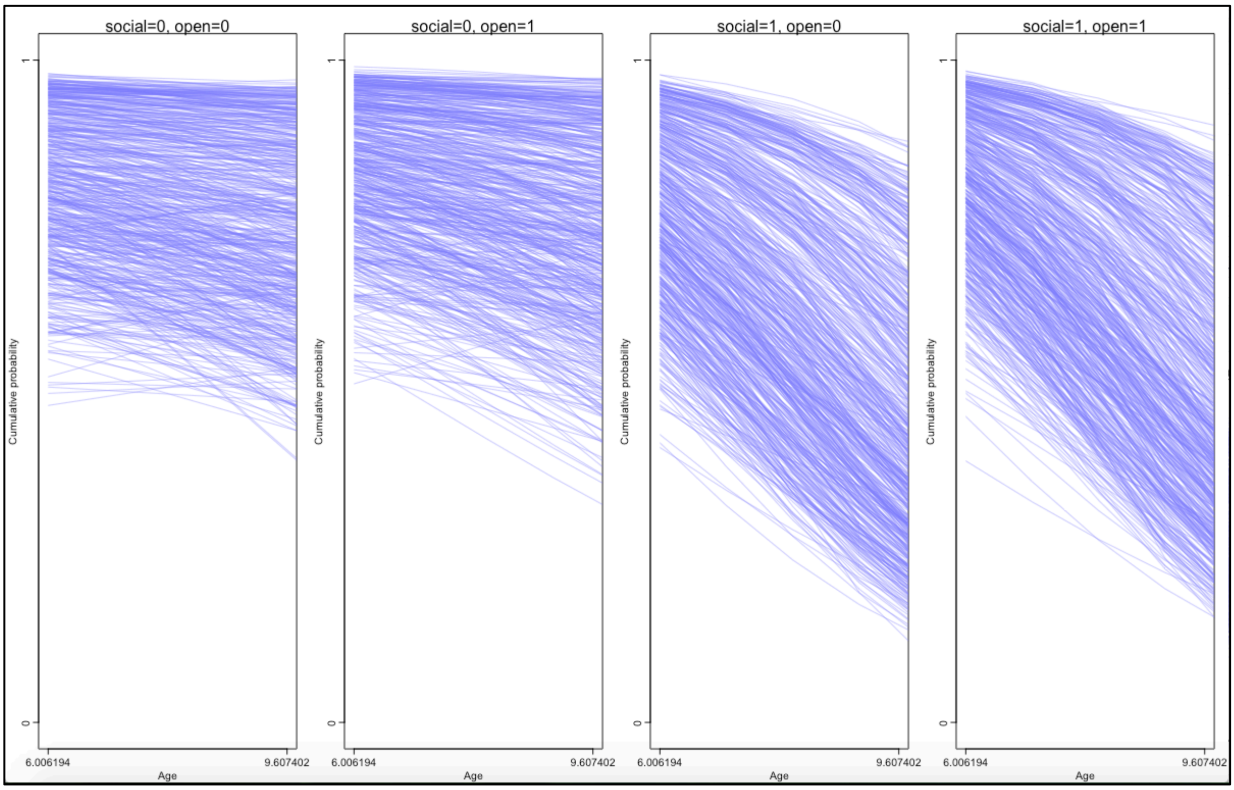

***Figure 5*. Four graphs (A, B, C, and D from left to right) illustrating Model 2's predicted effects of participant age on macrostructure similarity scores. The 'low' age was set one standard deviation below the mean age of the entire usable dataset (7.81), while the 'high' age was set one standard deviation above this mean. Graphs A and B, on the left, show the effect of increasing participant age when the model was asocial, while graphs C and D, on the right, show the effect of increasing participant age when the model was social. Graphs A and C, on the far left and second from right respectively, show the effect of increasing participant age when the task was close-ended, while graphs B and D, on the second from left and far right respectively, show the effect of increasing participant age when the task was open-ended. For all four graphs, the degree of internal evidence of failure of the participant was set at 0.69: one standard deviation below the mean score for the entire usable dataset (4.62).**

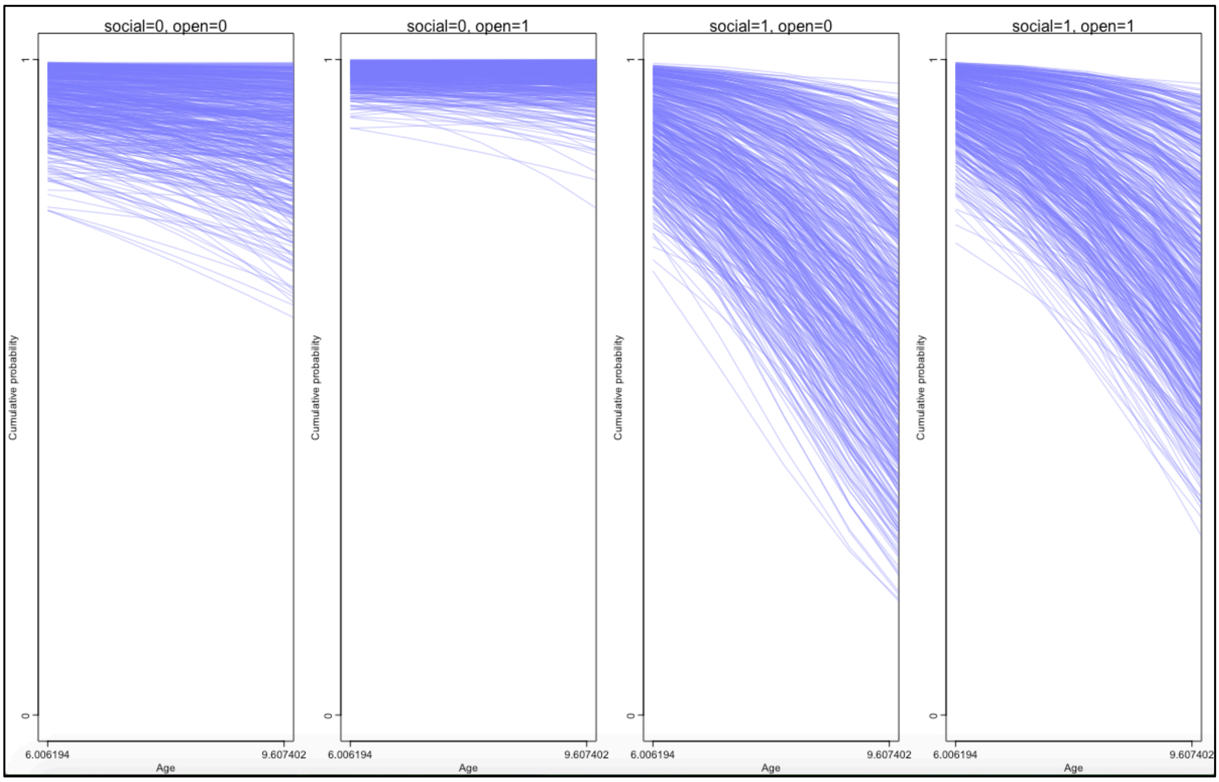

***Figure 6*. Four further graphs (A, B, C, and D from left to right) illustrating Model 2's predicted effects of increased participant age on macrostructure similarity scores. The 'low' age was set one standard deviation below the mean age of the entire usable dataset (7.81), while the 'high' age was set one standard deviation above this mean. The interactions described in the four graphs, between the asocial and social models, and close- and open-ended conditions, were the same as the graphs in Figure 5. However, for all of these graphs, the degree of internal evidence of failure of the participant was instead set at 8.54: one standard deviation above the mean score for the entire usable dataset (4.62).**

4.2: Participant sex

The tentative hypothesis that females would show increased microstructure and macrostructure similarity scores appears to have been supported in older but not younger children. Across eight microstructure and macrostructure graphs with a social model and older participant age (Figures 9, 10, 13, and 14), only one graph did not show a reliably positive effect of being female. Across eight microstructure and macrostructure graphs with an asocial model and older participant age (Figures 9, 10, 13, and 14), seven graphs did not show reliably positive effects of being female. This would be in line with the hypothesis. However, the effect of being female was more varied amongst younger children, across microstructure and macrostructure scores. This seems related to the findings above, which revealed the positive effect of age itself on microstructure and macrostructure similarity in this dataset.

The effect of being female, rather than male, often appears to have increased microstructure similarity scores. This was the case in 11 out of the 16 graphs between Figures 7, 8, 9, and 10. However, there were also five graphs in which the change from male to female seems to have had a negative effect on microstructure similarity. The effect of being female thus seems to have depended on other conditions which the participant was building under. The influence of the social model (graphs A and B across Figures 7 to 10), rather than the asocial model (graphs C and D across Figures 7 to 10), on the effect of a female participant on microstructure similarity appears to have been complex. There were examples of the social model reversing the effect of the 'female' variable on microstructure with the asocial model, as well as examples of the social model maintaining the effect of 'female' with an asocial model. It appears that the tentative hypothesis of females showing higher microstructure similarity scores than males, when the model is social, was upheld among older but not younger children. The hypothesis appears contradicted (i.e., where there was a social model females showed lower microstructure similarity scores than males) when participant age was low and the task was close-ended. Furthermore, where participant age was low and the task open-ended, the positive effect on microstructure similarity of being female with a social model was no different to that with an asocial model. More detailed description of the results of this statistical analysis, and the structure of the statistical model which generated it (Model 3), can be found in Appendix 6.3.

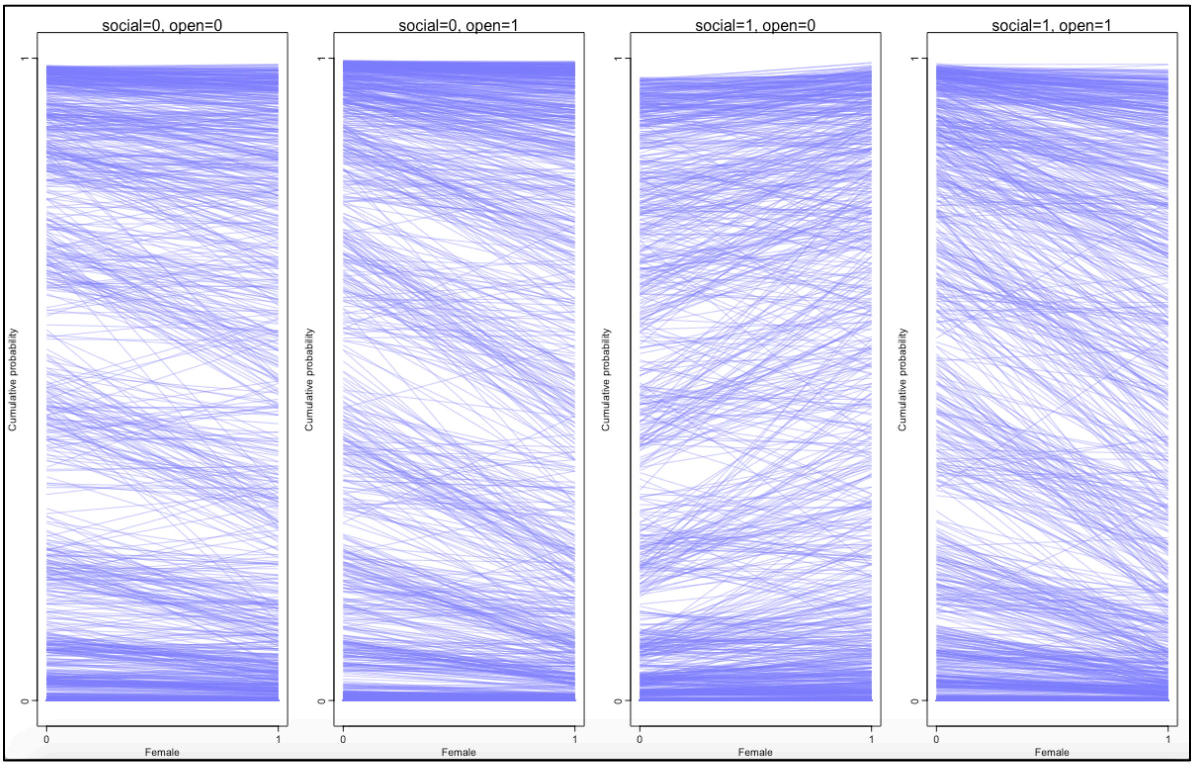

Figure 7. **Four graphs (A, B, C, and D from left to right) each showing Model 3's predicted effects of turning a male participant (left, Female=0) into a female participant (right, Female=1). Graphs A and B, on the left, show the result of this change when the model was asocial, rather than social as shown in graphs C and D on the right. Graphs A and C, on the far left and second from right respectively, show the effect of the variable 'female' when the task was close-ended, as opposed to open-ended as shown in graphs B and D on the second from left and far right respectively. All four graphs set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset) and degree of internal evidence of failure at 'low' too (i.e., 0.69, one standard deviation below the mean score of the entire usable dataset).**

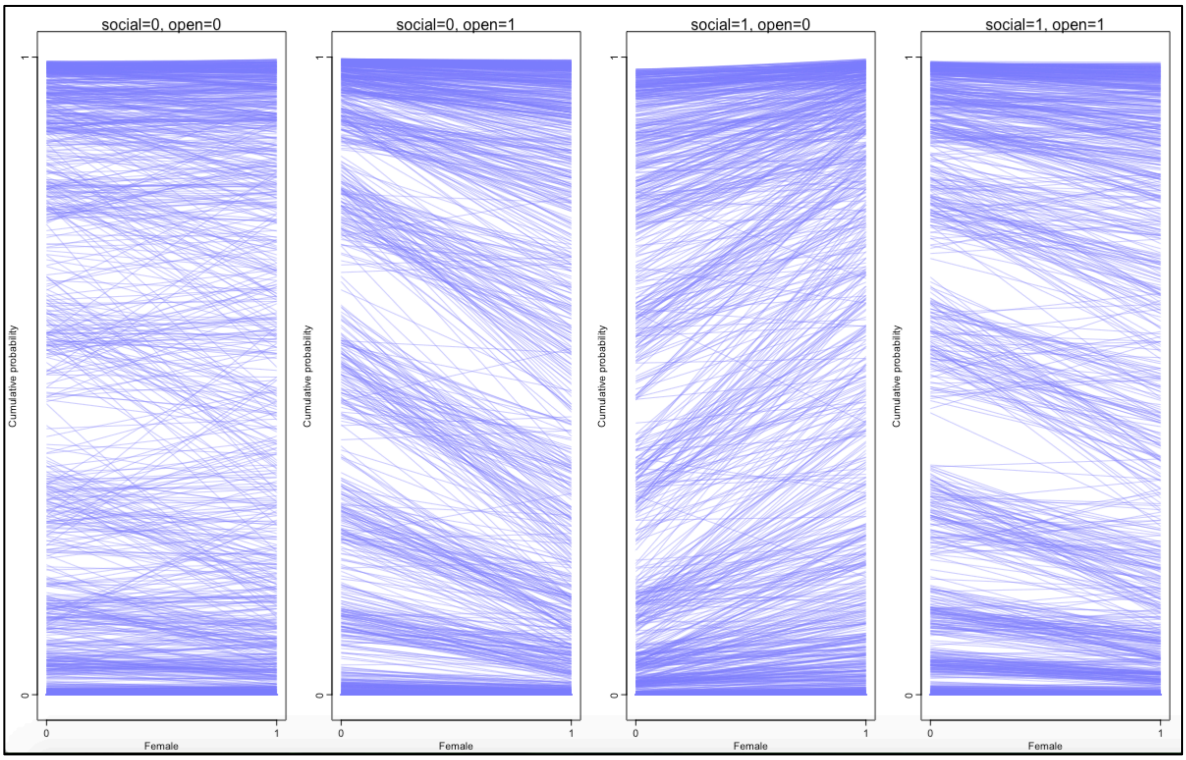

Figure 8. **Four graphs (from left to right: A, B, C, and D) each showing Model 3's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models, and close- and open-ended tasks were the same as in Figure 7. These four graphs also set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset). However, the degree of internal evidence of failure was set at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

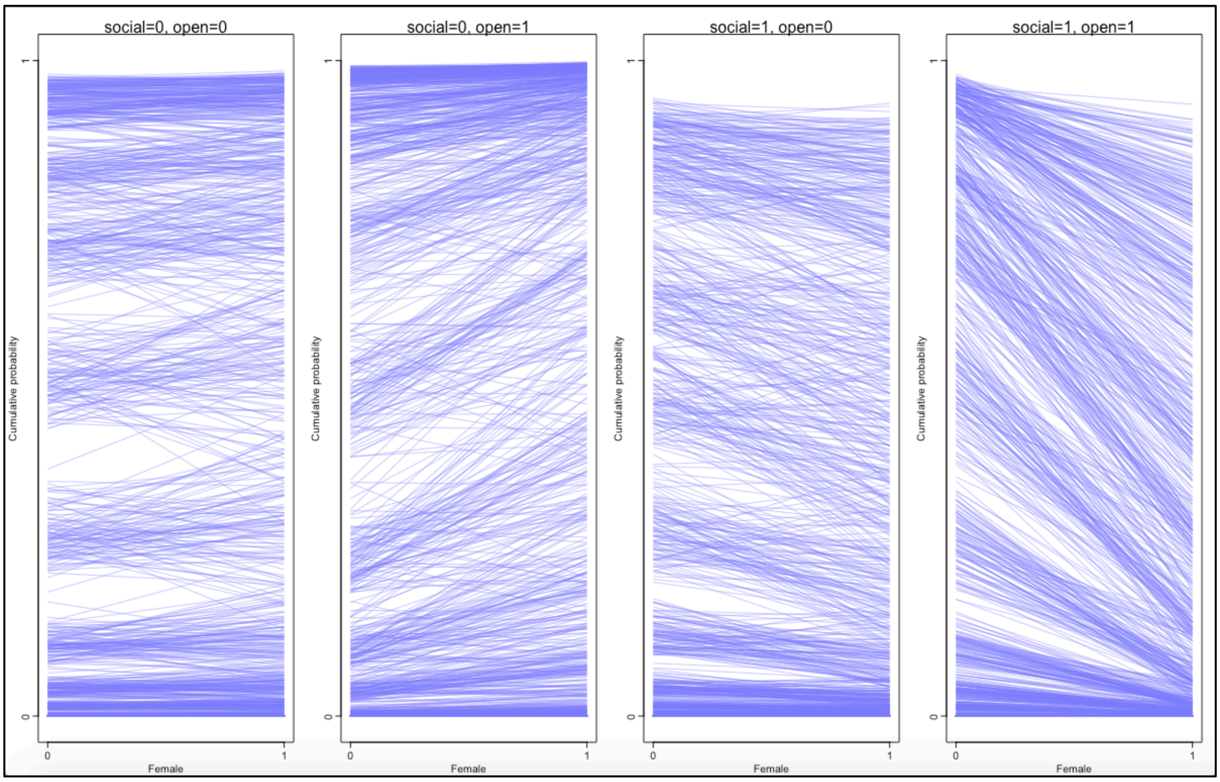

***Figure 9*. Four graphs (A, B, C, and D, from left to right) each showing Model 3's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models, and close- and open-ended tasks were the same as in Figure 7. These four graphs set participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset). The degree of internal evidence of failure was kept 'low' (i.e., 0.69, one standard deviation below the mean score of the entire usable dataset).**

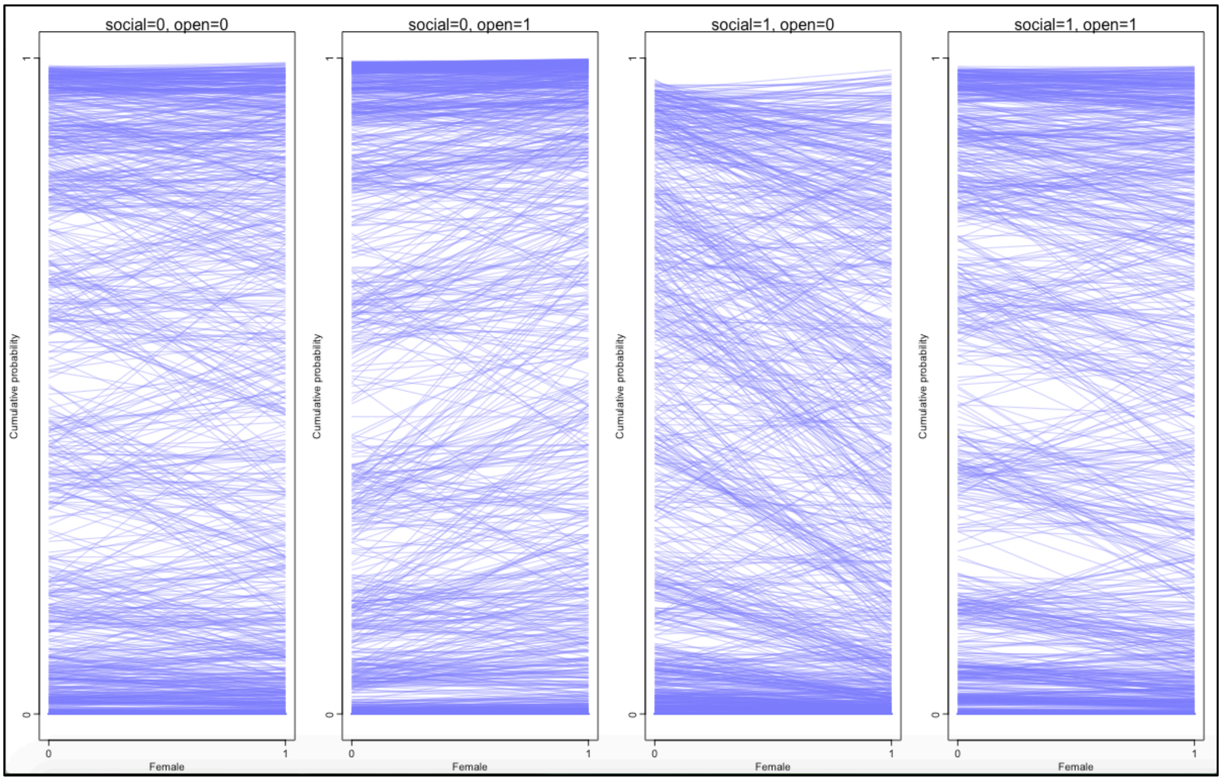

***Figure 10*. Four graphs (from left to right: A, B, C, and D) each showing Model 3's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models, and close- and open-ended tasks were the same as Figure 7. These four graphs set both participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset) as well as the degree of internal evidence of failure at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

It is noteworthy that of five graphs in which macrostructure was recognisably less similar to the social model in females than in males (graphs B and C in Figure 12, graph A in Figure 13, and graphs A and B in Figure 14), only one included a social model (which was Figure 12's graph C). This suggests that the social model made a negative effect of being female on microstructure similarity scores less likely. More often, with a social model, the effect of 'female' was to increase macrostructure similariy scores: in arguably 4 of the 8 graphs which did have a social model. However, this effect was not present in younger children exhibiting lower internal evidence of failure (in either the open- or close-ended task; Figure 11's graphs C and D), in younger children exhibiting higher internal evidence of failure with the close-ended task (as already discussed, Figure 12's graph C), or in older children exhibiting lower internal evidence of failure with a close-ended task (Figure 13's graph C). A more detailed description of the results of this statistical analysis, and the structure of the statistical model which generated it (Model 4), can be found in Appendix 6.4.

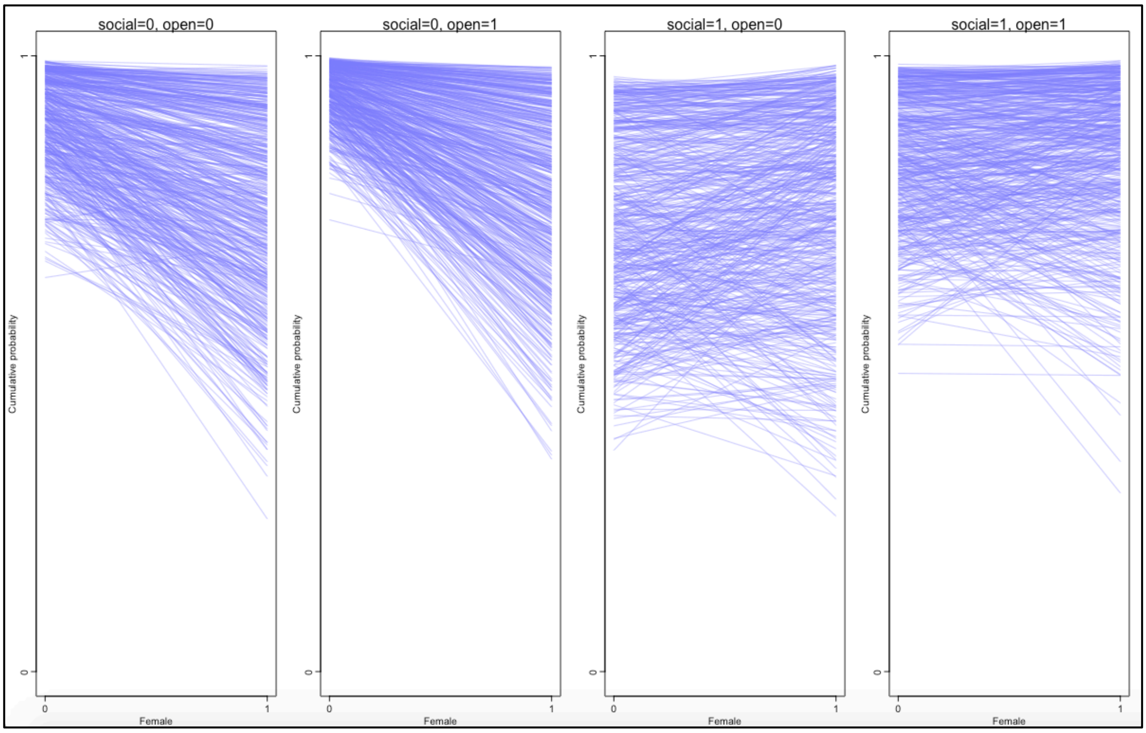

***Figure 11*. Four graphs (from left to right: A, B, C, and D) each showing Model 4's predicted effects of turning a male participant (left, Female=0) into a female participant (right, Female=1). Graphs A and B (on the left) show the result of this change when the model was asocial, rather than social as shown in graphs C and D (on the right). Graphs A and C (on the far left and second from right, respectively) show the effect of the variable 'female' when the task was close-ended, as opposed to open-ended as shown in graphs B and D (on the far right and second from left respectively). All four graphs set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset), and the degree of internal evidence of failure at 'low' too (i.e., 0.69, one standard deviation below the mean score of the entire usable dataset).**

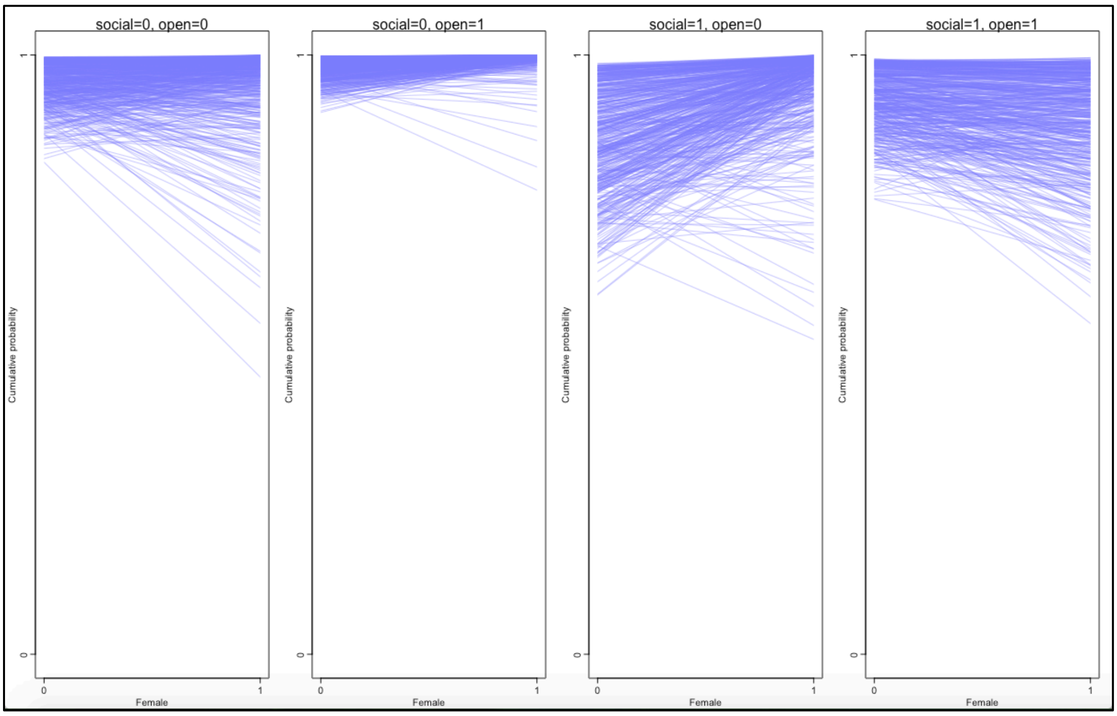

Figure 12. **Four graphs (from left to right: A, B, C, and D) each showing Model 4's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models (graphs A and B versus C and D, respectively), and the close- and open-ended task (graphs A and C versus B and D, respectively) were the same as in Figure 11. These four graphs also set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset). However, the degree of internal evidence of failure was set at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

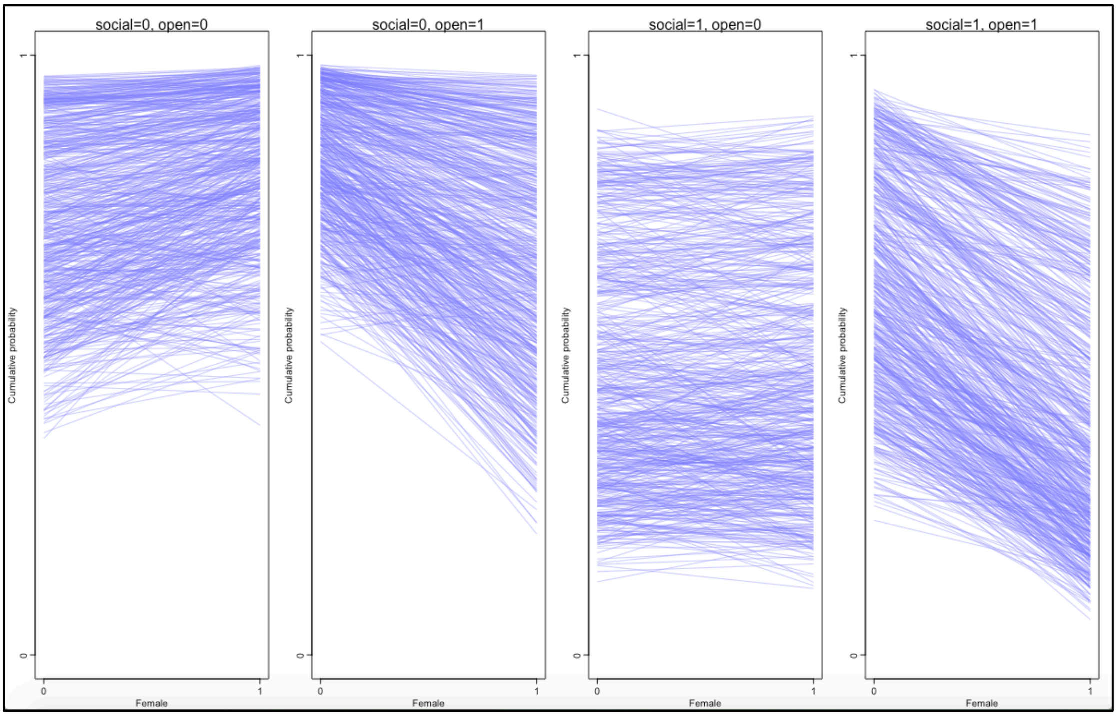

Figure 13. **Four graphs (from left to right: A, B, C, and D) each showing Model 4's predicted effects of turning a male participant into a female participant. The interactions between asocial and social models (graphs A and B versus C and D, respectively), and the close- and open-ended task (graphs A and C versus B and D, respectively) were the same as in Figure 11. These four graphs set participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset). The degree of internal evidence of failure was kept 'low' (i.e., 0.69, one standard deviation above the mean score of the entire usable dataset).**

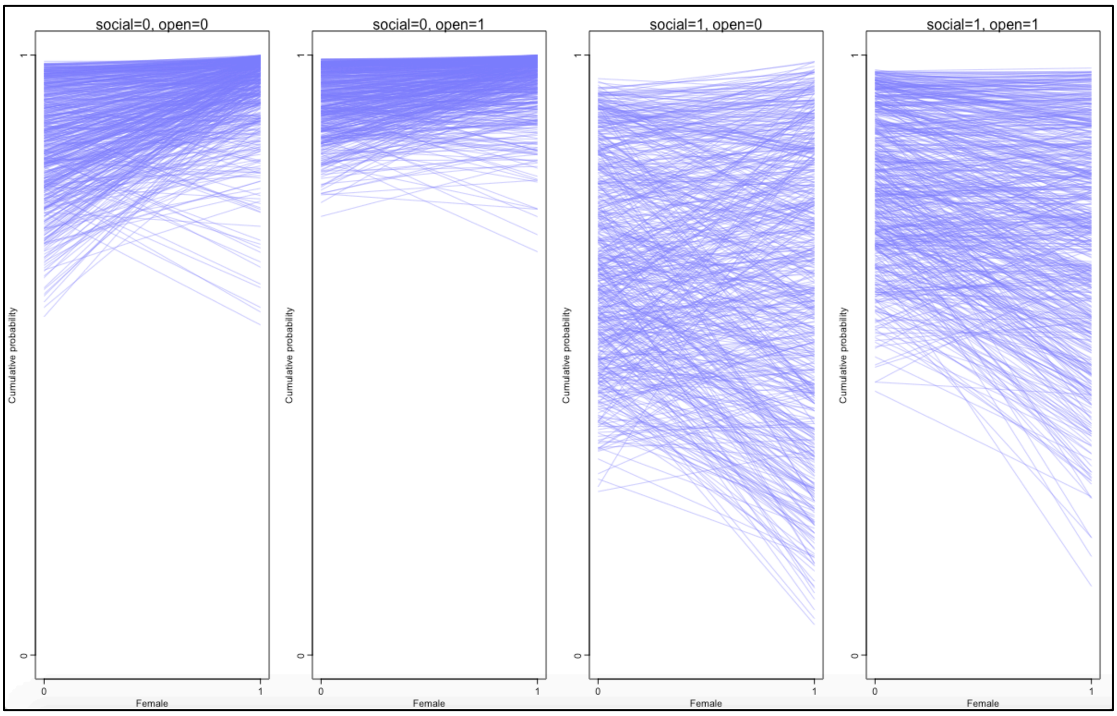

***Figure 14*. Four graphs (from left to right: A, B, C, and D) each showing Model 4's predicted effects of turning a male participant into a female participant. The interactions between asocial and social models (graphs A and B versus C and D, respectively), and the close- and open-ended task (graphs A and C versus B and D, respectively) were the same as in Figure 11. These four graphs set both participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset) as well as the degree of internal evidence of failure at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

4.3: Participant attendance to the video

The 'attendance to the video' variable did not appear to conform to the hypothesis that it would be associated with higher microstructure and macrostructure similarity scores. Participants' attendance to the experimental video did appear to have directional relationships with the two outcome variables; most of the relationships are not null or insignificant. However these relationships varied widely between different conditions such that any consistent effect of increased attendance to the video was lost. This was true also for its interactions with the social versus asocial model. The data thus indicated that attendance to the video was generally not an important factor in creating the similarity of a participant's microstructure or macrostructure design via copying behaviour.

In interaction with three other variables, the effect of higher attendance to the video on microstructure similarity scores appears to have been positive in just two out of eight conditions graphed: with the asocial model and close-ended task among younger children (graph A, Figure 15), and the social model and close-ended task among older children (graph C, Figure 16). There appears to have been a somewhat negative relationship between attendance scores and microstructure similarity in four out of the eight conditions: with the asocial model and open-ended task (graph B), the social model and close-ended task (graph C), the social model and open-ended (graph D) in Figure 15, and the asocial model and open-ended task (graph B) in Figure 16. In most of these six cases, any effect attendance to the video has on microstructure similarity was weak. The two conditions where any effect was strongest were found in Figure 16 (i.e., with older children): with the asocial model and open-ended task (graph B), and with the social model and close-ended task (graph C). In two of the

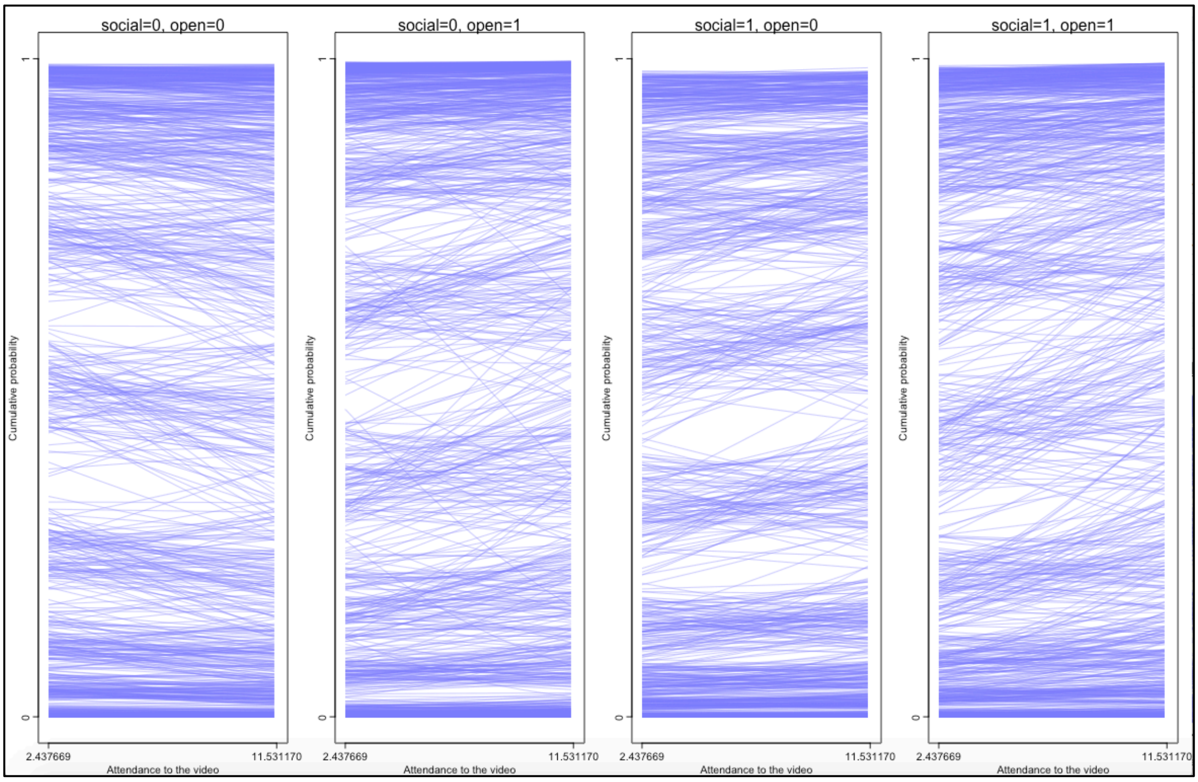

***Figure 15*. Four graphs (from left to right: A, B, C, and D) describing Model 5's predictions for the effect on microstructure similarity scores of turning low participant attendance to the experimental video into high participant attendance to the experimental video. Graphs A and B (on the left) show the result of this change when the model was asocial, rather than social as shown in graphs C and D (on the right). Graphs A and C (on the far left and second from right) show the effect of the variable 'attendance score' when the task was close-ended, as opposed to open-ended as shown in graphs B and D (on the second from left and far right). All four graphs set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset).**

conditions with older children (i.e., in Figure 16) there appears to have been a neutral effect of turning low attendance into high attendance: with the asocial model in the close-ended task (graph A) and with the social model in the open-ended task (graph D). The effect of attendance to the video was thus mostly weak, more often negative on microstructure scores than positive, though with variation in its effects across different conditions. More detailed description of the results of this statistical analysis, and the structure of the statistical model which generated it (Model 5), can be found in Appendix 6.5.

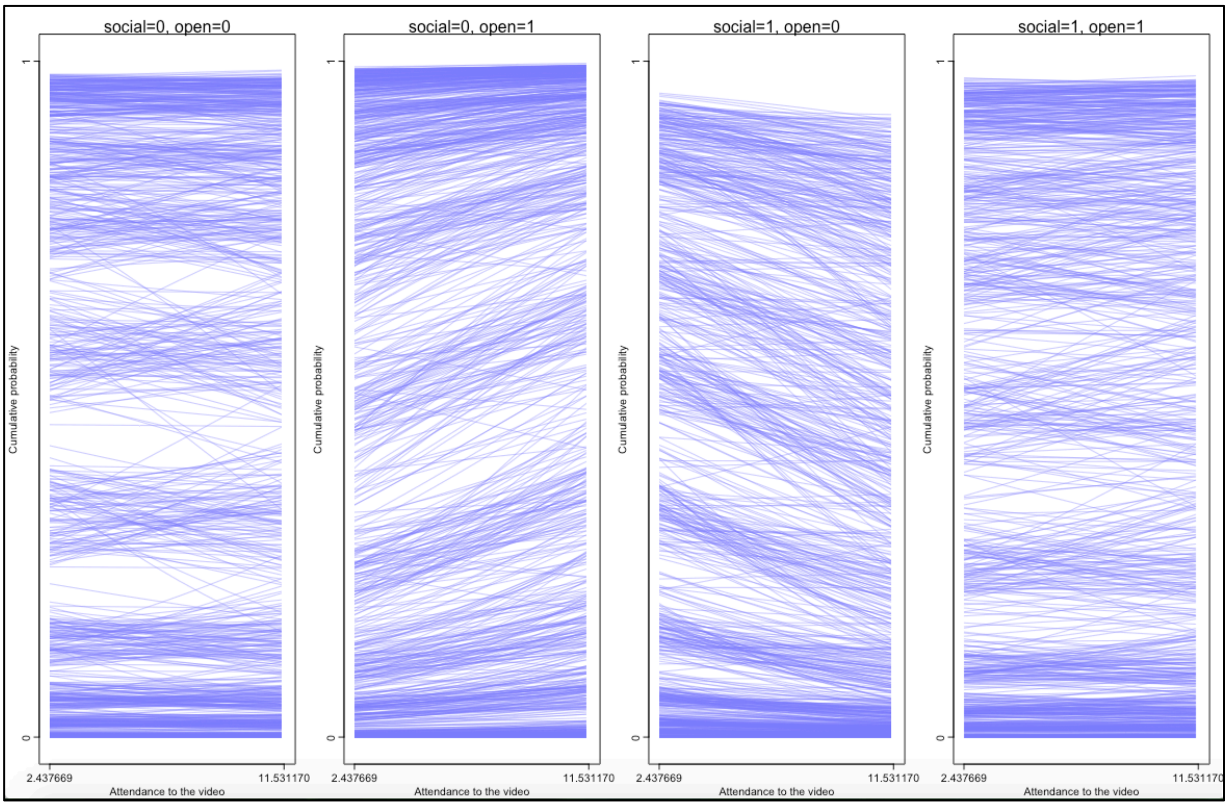

***Figure 16*. Four graphs (from left to right: A, B, C, and D) describing Model 5's predictions for the effect on microstructure similarity scores of turning low participant attendance to the experimental video into high participant attendance to the experimental video. The interactions between social (graphs C and D) and asocial models (graphs C and D), and close-ended (graphs A and C) and open-ended tasks (graphs B and D) were the same as Figure 15 above. These four graphs set participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset).**

The effect of participants' attendance to the video on macrostructure similarity scores appears to have varied across conditions. Across most conditions, the change from low to high attendance to the video had little predicted effect on macrostructure similarity. It had a clearly positive effect on macrostructure similarity scores in perhaps five out of the sixteen conditions graphed: graph A in Figure 17 (i.e., younger children exhibiting lower internal evidence of failure), graphs A, C, and D in Figure 18 (younger children exhibiting higher internal evidence of failure), and graph C in Figure 20 (older children exhibiting higher internal evidence of failure). It had a somewhat negative effect on macrostructure similarity in another seven of the conditions: graphs C and D in Figure 17, graph B in Figure 18, graphs A and B in Figure 19, and graphs A and B in Figure 20. However, the real effects of this negative relationship on macrostructure scores were small. This leaves 5 of the 16 conditions in which there seems to have been no particularly directional effect of attendance scores on macrostructure similarity scores. This indicates that the effect of 'attendance to the video' was dependent on other factors, including the experimental condition, and that across most conditions it was not a particularly strong influence on macrostructure similarity score variation. See Appendix 6.6 for further detail of the results presented between Figures 17 and 20, and the structure of the statistical model which generated it (Model 6).

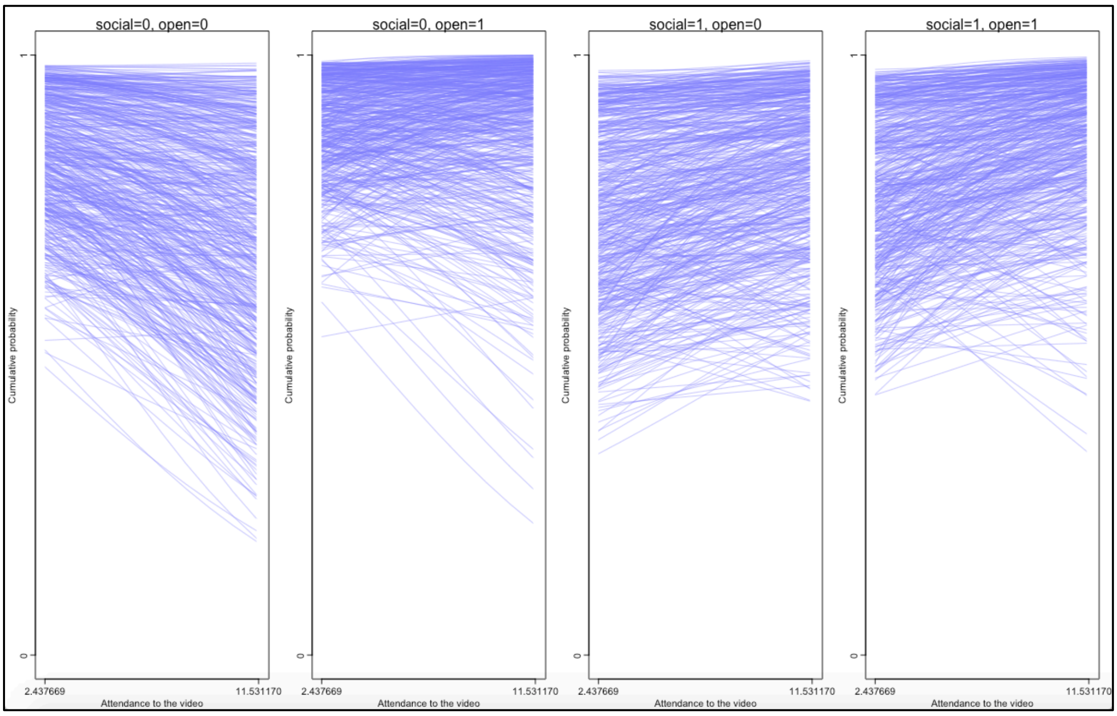

***Figure 17*. Four graphs (from left to right: A, B, C, and D) each showing Model 6's predicted effects of turning a participant with low attendance to the video (left) into one with high attendance (right). Graphs A and B (on the left) show the result of this change when the model was asocial, rather than social as shown in graphs C and D (on the right). Graphs A and C (on the far left and second from right) show the effect of the variable 'attendance score' when the task was close-ended, as opposed to open-ended as shown in graphs B and D (on the second from left and far right). All four graphs set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset), and degree of internal evidence of failure at 'low' too (i.e., 0.69, one standard deviation below the mean score of the entire usable dataset).**

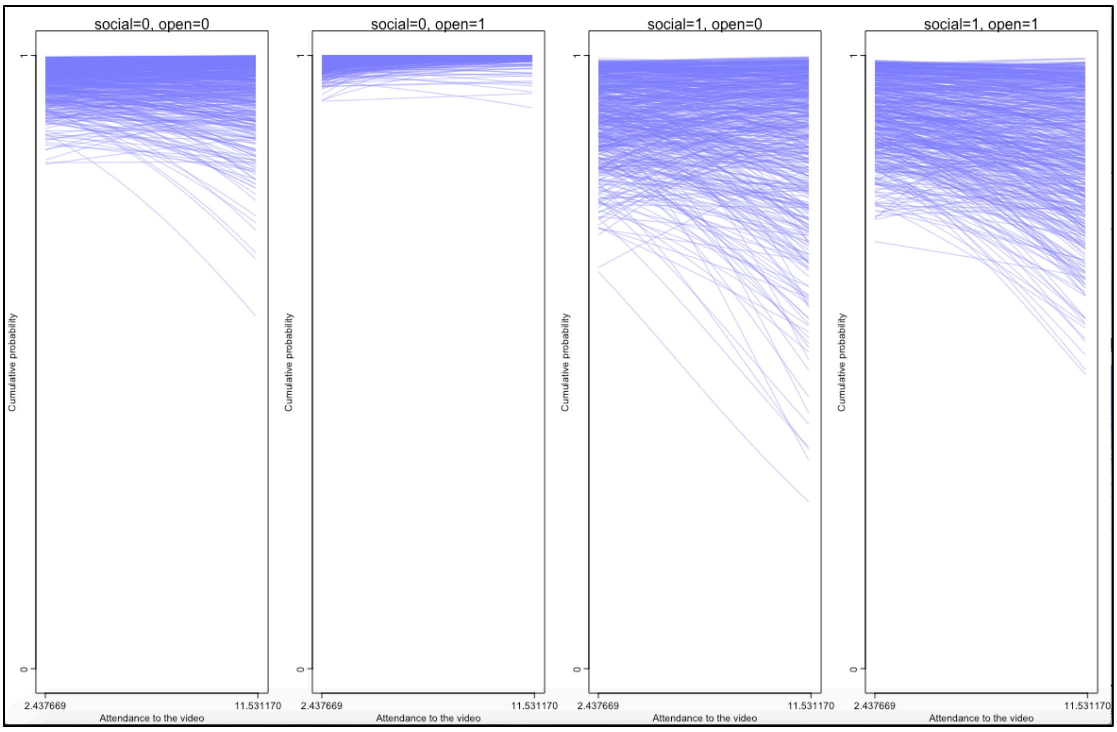

Figure 18. **Four graphs (from left to right: A, B, C, and D) each showing Model 6's predicted effects of turning a male participant into a female participant. The interactions between social (graphs C and D) and asocial models (graphs A and B), and close-ended (graphs A and C) and open-ended tasks (graphs B and D) were the same as in Figure 17. These four graphs also set participant age at 'low' (i.e., 6.01, one standard deviation below the mean age of the entire usable dataset). However, the degree of internal evidence of failure was set at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

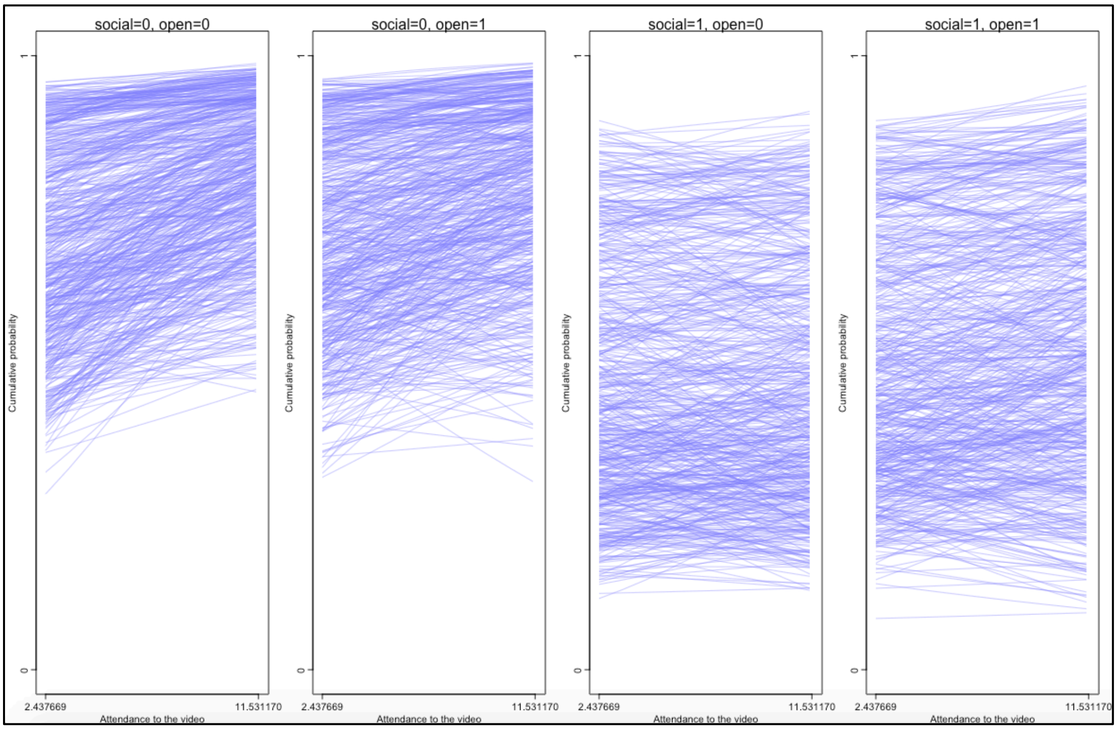

Figure 19. **Four graphs each showing Model 6's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models, and close- and open-ended conditions were the same as Figure 17. These four graphs set participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset). The degree of internal evidence of failure was kept 'low' (i.e., 0.69, one standard deviation below the mean score of the entire usable dataset).**

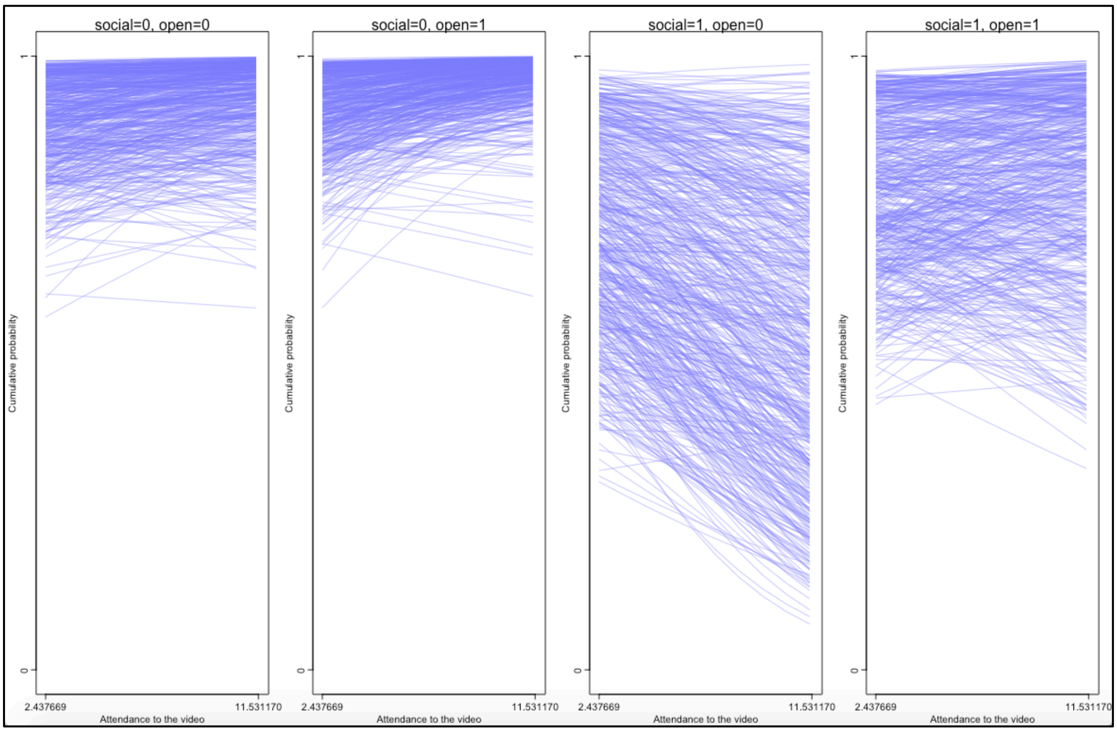

***Figure 20*. Four graphs each showing Model 6's predicted effects of turning a male participant into a female participant. The interactions between social and asocial models, and close- and open-ended conditions were the same as Figure 17. These four graphs set both participant age at 'high' (9.61, one standard deviation above the mean age of the entire usable dataset) as well as the degree of internal evidence of failure at 'high' (i.e., 8.54, one standard deviation above the mean score of the entire usable dataset).**

Chapter 5: Results and discussion of the effects of the close-ended task

In Chapter 5 I ask: 'How is microstructure and macrostructure copying affected by close-ended conditions as used in previous social learning studies?' This question was addressed by statistically comparing the similarity of children's buildings across the eight conditions outlined in Chapter 3. I then discuss the results for each of the four hypotheses in order to make sense of the information presented in terms of its implications for the current thesis. In broad terms, I argue data indicated that (1) in the close-ended task there was greater similarity of macrostructure designs to the social model than in the open-ended task, (2) that in close-ended conditions there was greater microstructure similarity to the social model when the model was social rather than asocial, and (3) that in close-ended conditions the successful social model (rather than the unsuccessful social model) caused greater microstructure similarity to the social model. However, I also argue (4) that data gave only the weakest support, among younger children, to the prediction that social model success would not be a useful predictor of macrostructure similarity in the close-ended task. Furthermore, a predominant theme of the results was variation and interdependence. The hypothesised effects of the main predictor variables were, in most cases, supported by data; yet these effects varied in strength, as well as sometimes direction, dependent on the status of other variables included in each model.

5.1: Hypothesis 1

In Hypothesis 1 I predicted that, across all conditions, the close-ended task (versus the open-ended task) would be a positive predictor of macrostructure similarity scores, since only in close-ended conditions were participants verbally instructed to build the tallest tower. The model for this first hypothesis had access to 559 cases, with the 6 cases removed being the same as those above for the macrostructure data in Chapter 4. As for Chapter 4, I compared models with many permutations of interactions between different predictor variables. Descriptions of these various models, and an account of the comparisons between the models, can be found in Appendix 7.1.

(Model 7)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_{CS} CS_i + \beta_{CN} CN_i + \beta_{CG} CG_i + \beta_{CSN} CSN_i + \beta_{CSG} CSG_i + \beta_{CNG} CNG_i + \beta_{CSNG} CSNG_i$$

The main predictor is the variable ‘close’ (C, as opposed to O, the variable used in Chapter 4 to denote the presence of the open-ended task), which interacted with the other predictor variables: the social rather than asocial model (S), the degree of the participant’s internal evidence of failure (N), and the age of the participant (G). Adding interactions with any more variables, such as the sex of the participant or the success of the model, or swapping these variables for the ones in the above model, resulted in worse estimates. I look first at the posterior mean for ‘close’. The influence of ‘close’ on the outcome variable appears to have been strong: its mean posterior main effect was 2.36. However, large variation in the posterior distribution (SD=1.52; HDPI=0.89, between -0.05 and 4.76) indicates that the effect of the close-ended task may have been dependent on other sources of variation in the model. The marginal effect of ‘close’, as calculated by Model 7, is illustrated in Figure 21. Since Model 7 had four variables, each interacting with each other in various combinations, it is necessary to use graphs to understand how the effects of each of them are dependent on the effects of the others. While Figure 21 illustrates the marginal effect of the ‘close’ variable, to grasp the real effect of ‘close’ on the outcome variable it must be considered at the same time as the effects of the other predictor variables in Model 7.

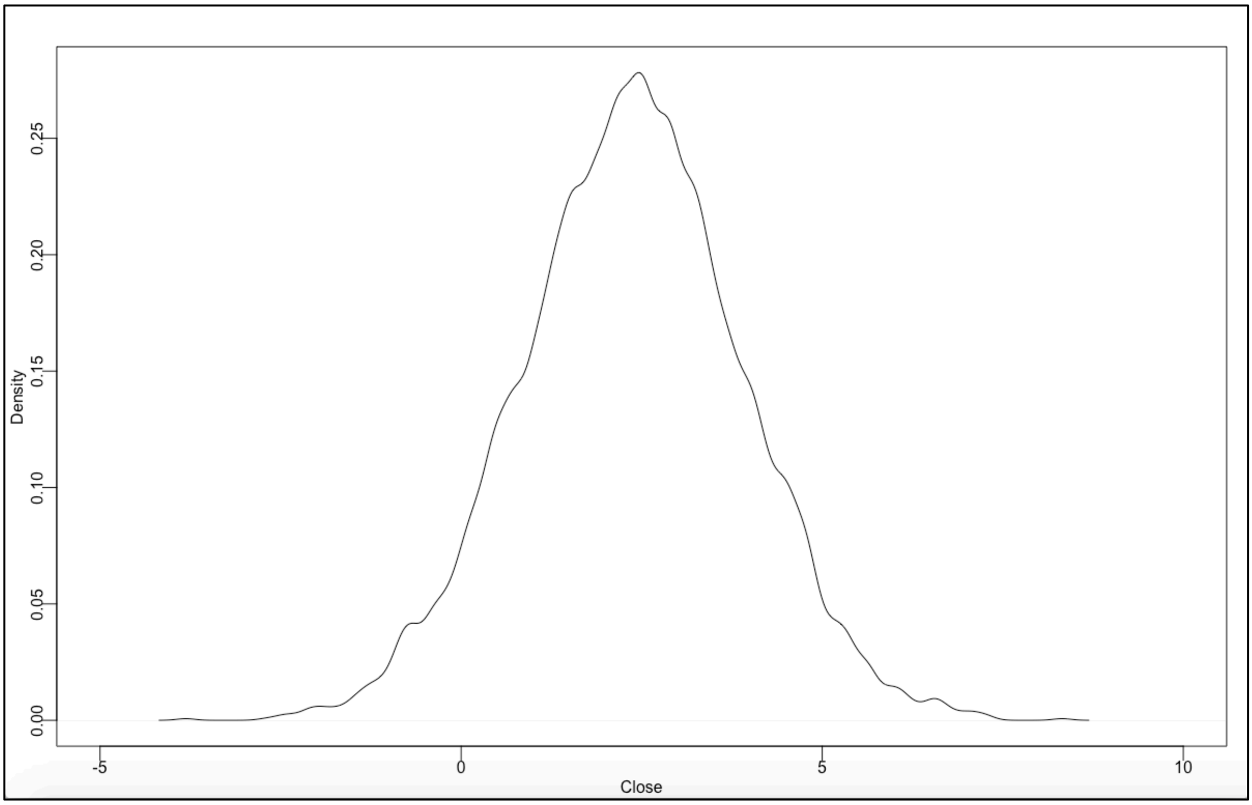

***Figure 21*. 1000 samples from the posterior distribution for the marginal effect of 'close', produced by Model 7. This graph shows the range of values (-5 to positive 10, on the horizontal axis) which the effect of 'close' could have on the outcome variable, and the posterior probability (on the vertical asxis) that Model 7 assigned to these values.**

Figures 22 and 23 show the predicted effects of close-ended conditions on macrostructure similarity scores, when participants observed either a social or asocial model, experienced either a high or low degree of internal evidence of failure, and were either younger or older. Under a number of conditions, the predicted effect of 'close' was to increase macrostructure similarity scores. This was consistent with the indications of Figure 21 and Hypothesis 1. However, this positive effect of 'close' on macrostructure similarity varied with the influence of other variables. There was a clearly positive effect of 'close' in at least four of the eight conditions: (1) the asocial model and low internal evidence of failure (graph A, Figure 22) and (2) the asocial model and high internal failure in younger children (graph B, Figure 22), and (3) the asocial model and high internal failure (graph B, Figure 23) and (4) the social model and high internal failure in older children (graph D, Figure 23).

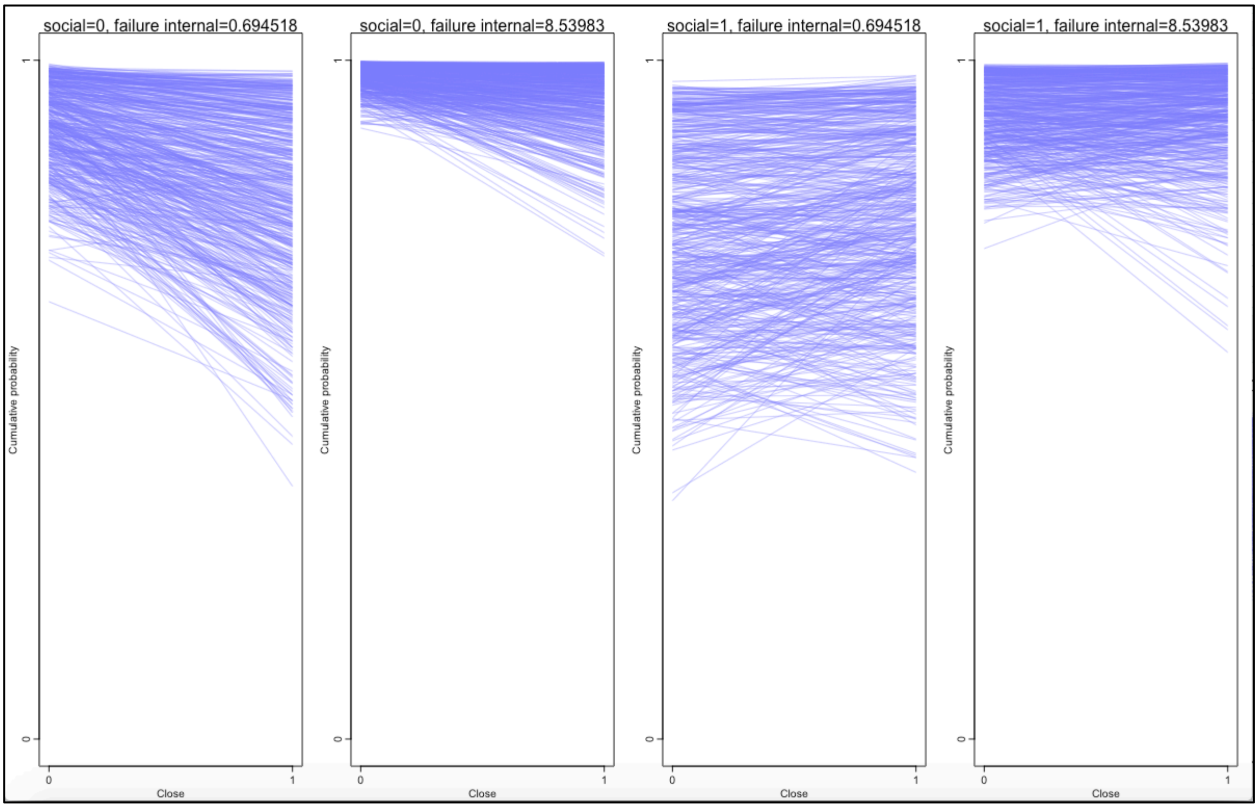

***Figure 22*. Four graphs (from left to right: A, B, C, and D) describing Model 7's predicted effects of 'close' on macrostructure similarity scores. Graphs A and B (on the left) show the effect of 'close' with a model irrelevant for children's building (i.e. the 'asocial' model), while graphs C and D (on the right) show the effect of 'close' with a social model. Graphs A and C (on the far left and the second from right) describe the effect of 'close' with low internal evidence of failure. 'Low' internal evidence of failure was set at one standard deviation below the mean of 4.62. Graphs B and D (the second from left and far right) show the effect of 'close' with high internal evidence of failure: one standard deviation above the mean. The participants' age was, in these four graphs, set at 6.01 years old. This was one standard deviation below the mean age of the participants across all conditions (7.8 years old).**

This is a less consistent effect than what was predicted in Hypothesis 1. Interestingly, the two conditions in which 'close' seems to have had the greatest effect were opposite of one another: younger children in asocial conditions with low internal evidence of failure (graph A, Figure 22), and older children in social conditions with high evidence of failure (graph D, Figure 23). The effect of 'close' among the younger children with the asocial model and low internal evidence of failure (graph A, Figure 22) disappeared for older children in the equivalent condition (graph A, Figure 23). The effect of 'close' in older children with a social model and high internal evidence of failure (graph D, Figure 23) appeared to diminish with younger children (graph D, Figure 22) too. Two conditions appeared to show no real change at all in macrostructure similarity scores. These two conditions were also quite different from one other: in younger children with a social model and low internal evidence of failure (graph C, Figure 22), and in older children with an asocial model and also low internal evidence of failure (graph A, Figure 23). The hypothesis was thus generally supported, but with exceptions. See Appendix 8.1 for some further detail of the results presented in Figures 22 and 23.

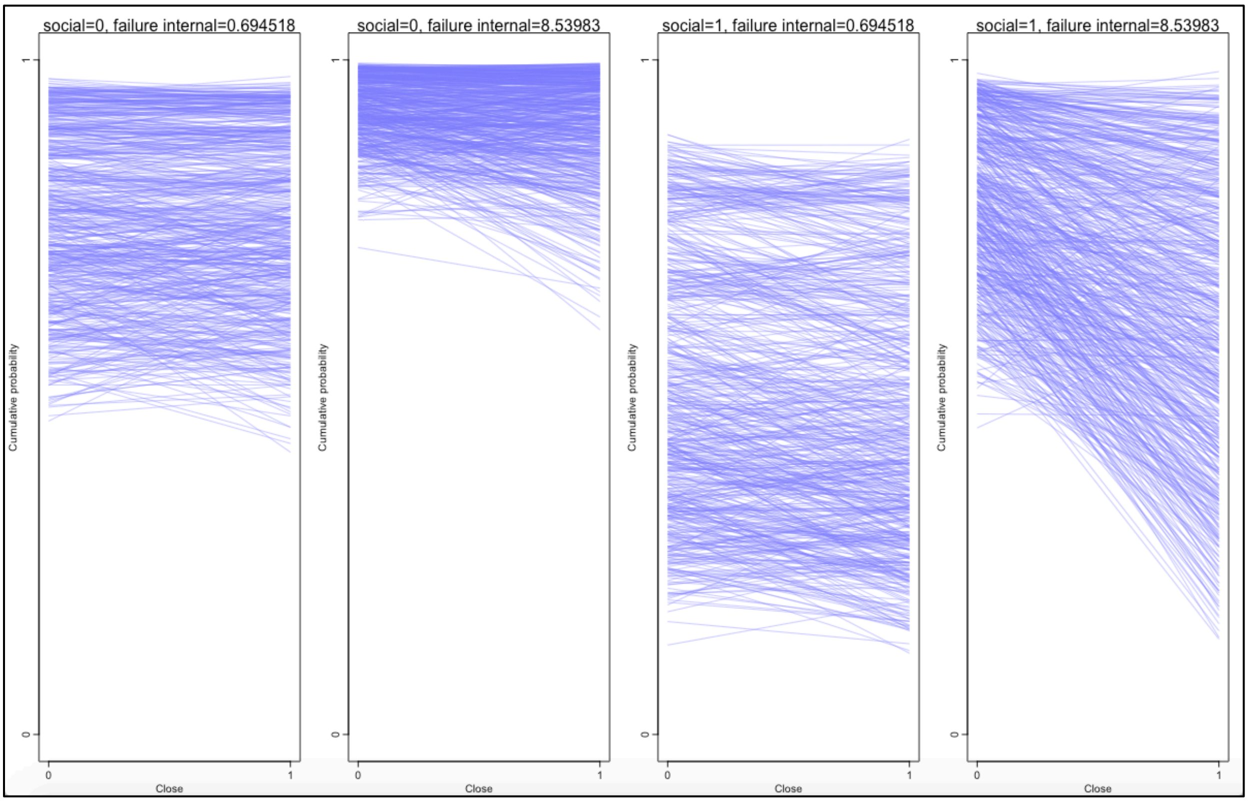

***Figure 23*. Four graphs (from left to right: A, B, C, and D) showing the effects of 'close' on macrostructure similarity. The interactions between the social model (graphs C and D) versus asocial model (graphs A and B) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) were the same as Figure 22, except that in these four graphs, the children were 9.6 years old, one standard deviation above the mean age for all usable samples (7.8 years old).**

I now move on to discuss the implications of these results for the current study. Hypothesis 1 concerned how macrostructure copying varied between the close- and open-ended conditions. It predicted that across all conditions the close-ended task (versus the open-ended task) should be a positive predictor of macrostructure similarity scores. It justified this by arguing that only in close-ended conditions were participants verbally instructed to build the tallest tower, so they should be more constrained to building something akin to a tower than children who were not given these instructions. This hypothesis seems to have found some support in the data, as nearly all conditions showed at least a weakly positive effect of the close-ended task on macrostructure similarity. This means that Model 7 predicted, through its encounters with the data, that by exposing a participant to the close-ended task their build would likely exhibit greater macrostructure similarity to the model than if they had been exposed to the open-ended task. This is in line with the argument made above, that close-ended tasks restrict participants' building (see Carr 2016; Legare et al. 2015 for examples of the effectiveness of such verbal instruction).

It is important that Model 7 should predict that the positive effect would not be restricted to conditions with a social model, since the verbal instructions to participants may merely incentivise copying of the social model, thus having a lesser effect when the model is irrelevant to children's building (i.e., when the model is 'asocial'). Indeed Model 7 indicated that three out of four conditions with an asocial model showed a clearly positive effect of close-ended conditions on macrostructure similarity. However, only half of the eight conditions in total showed that this positive relationship was strong enough for the close-ended task to have had a meaningful impact on macrostructure similarity scores. Thus, the effect of the close-ended task on macrostructure similarity did seem dependent on the effects of other variables in interesting ways.

Three of the four conditions in which a positive effect of the close-ended task was harder to observe were found when participants observed the social model. One explanation could be that the effect of the social model increased macrostructure similarity to such a degree in the open-ended task that changing the task into a close-ended one had little effect on increasing macrostructure similarity further. This is given support by observing how much further the blue lines spread towards the bottom of the left hand side of graphs C and D in Figures 22 and 23, when the model was social, rather than asocial as in graphs A and B (respectively), and even more so when the participants were older and internal evidence of failure was lower. This would not explain, however, the lack of effect of the close-ended task on macrostructure similarity amongst older children exhibiting low internal evidence of failure and observing the asocial model (Figure 23's graph A). Furthermore, it does not explain how maybe the strongest effect of the close-ended task was found with older children exhibiting higher internal evidence of failure and with the social model (Figure 23's graph D). Perhaps the encounter with internal evidence of failure led participants to copy the model more closely, and stick to their instructions more rigidly. This would be in line with previous evidence that greater uncertainty in a task incentivises greater reliance on social information (Wood, Kendal & Flynn 2013a). In fact, higher internal evidence of failure seems predicted to increase the positive effect of the close-ended task in all conditions, including with the asocial model. The inverse interpretation thus appears more plausible: that participants engaging with the instructions to build a tower perhaps encountered less internal evidence of failure than participants building other, more 'experimental' structures. However, this would not explain why the effect of the close-ended task with younger children did not show the same influence of high versus low internal evidence of failure. Overall, a positive effect of the close-ended task on macrostructure similarity appears present under several conditions, but it was also influenced by other variables which could severely dampen its effect. Thus Hypothesis 1 found support particularly where the model was asocial (especially so in younger children) and where participant internal evidence of failure was high (especially so with older children).

5.2: Hypothesis 2

In Hypothesis 2 I predicted that, across the close-ended conditions, the presence of the 'social' model would be a positive predictor of variation in microstructure similarity scores, with higher scores than when participants observed the 'asocial' model. The data used here numbered 273 cases. This was due to the four cases dropped for the reasons discussed above, and the nature of the hypothesis, which did not make a prediction about the effects of social versus asocial models in open-ended conditions. The model comparison procedure was also undertaken for this hypothesis, an account of which can be found in Appendix 7.2. The description of the resulting model is:

(Model 8)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_{SU} SU_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{SNG} SUN_i + \beta_{SUN} SUG_i + \beta_{SUG} SNG_i + \beta_{SUNG} SUNG_i$$

In this model, the main predictor variable was 'social' (S). This variable interacted with three others: model success (U), internal (i.e. participant) evidence of failure (N), and participant age (G). The marginal effect of 'social', as calculated in Model 8, is illustrated in Figure 24. This first evidence appears to have contradicted Hypothesis 2, with the greater posterior probability lying over the negative values. Indeed the mean effect of 'social' was -1.39. However, there was posterior probability for a positive effect of 'social' as well. The mean effect had a large standard deviation of 1.71, and the 0.89 HPDI reached from -4.26 to positive 1.19. This indicates that the effect of 'social' may have been dependent on the effects of other variables. To see whether there was a difference between the effects of 'social' in interaction with these other variables, more graphs are required.

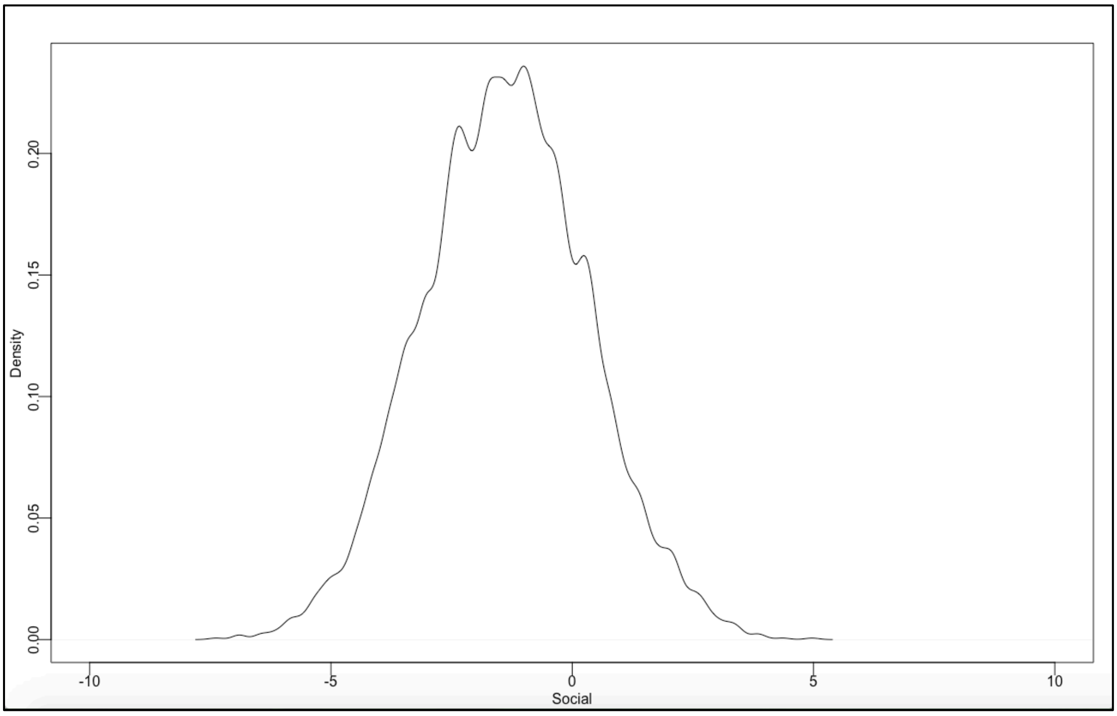

***Figure 24*. Graph showing the posterior distribution for the marginal effect of the variable 'social' on microstructure similarity scores.**

Figures 25 and 26 illustrate Model 8's predictions for the effect of a social model, in contrast to an asocial model, on microstructure similarity scores. Hypothesis 2's prediction of a positive effect of 'social' on microstructure similarity was clearly supported in six out of the eight conditions graphed below. There thus appear to have been two exceptions to the prediction. In young children with unsuccessful models and high internal evidence of failure (graph B, Figure 25), the social model appears to have decreased microstructure similarity scores. In young children with unsuccessful models and low internal evidence of failure (graph A, Figure 25), any positive effect of the social model appears to have been small. With these two concessions, therefore, Hypothesis 2 appears to have been supported.

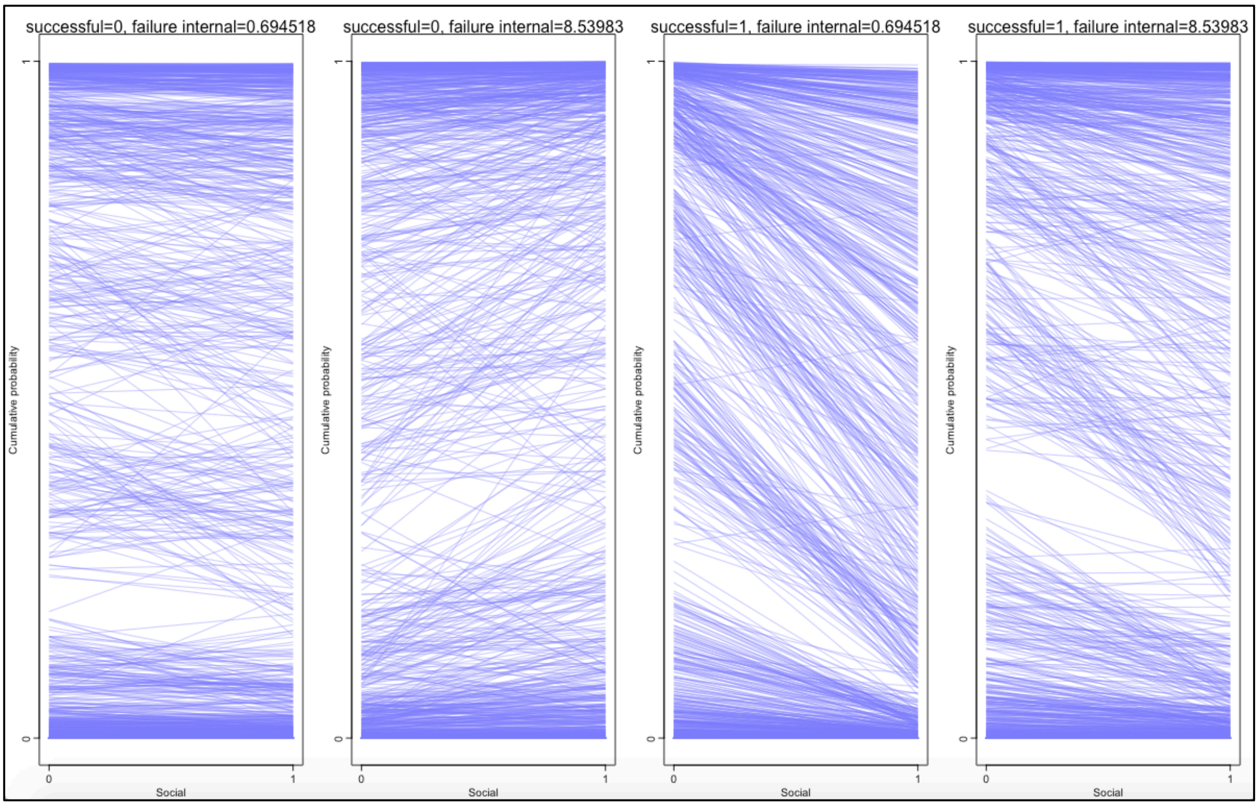

***Figure 25*. Four graphs (from left to right: A, B, C, and D) showing Model 8's predicted effects of changing an asocial model into a social model on microstructure similarity scores. The effect of a social model was here in interaction with the success of the model, and participants' internal evidence of failure and age. Graphs A and B (on the left) show the impact of a social model when the models (both social and asocial) were unsuccessful, while graphs C and D (on the right) show the impact of a social model when the models were successful. Graphs A and C (on the far left and second from right) show the impact of a social model when a participant's internal evidence of failure was one standard deviation below the mean score, while graphs B and D (second from left and on the far right) show the impact of a social model when a participant's internal evidence of failure was one standard deviation above the mean. For all four graphs, the participant's age was set at 6.01, one standard deviation below the mean of the entire usable dataset.**

In Figure 25, with younger children, a successful (graphs C and D) rather than unsuccessful social model (graphs A and B) was predicted to cause greater microstructure similarity scores with both low and high internal evidence of failure. With older children, in Figure 26, the effect of a successful social model (graphs C and D) made the already positive effect of an unsuccessful social model even stronger, across both low and high internal evidence of failure. When younger participants were more unsuccessful (graphs B and D in Figure 25) the effect of a social model on microstructure similarity was reduced compared to when they demonstrated less internal evidence of failure (graphs A and C). The effect of increasing participant age appears to have strengthened the positive effect of a social model across all conditions (graphs A to D between Figures 25 and 26). More detailed description of the results of this statistical analysis can be found in Appendix 8.2.

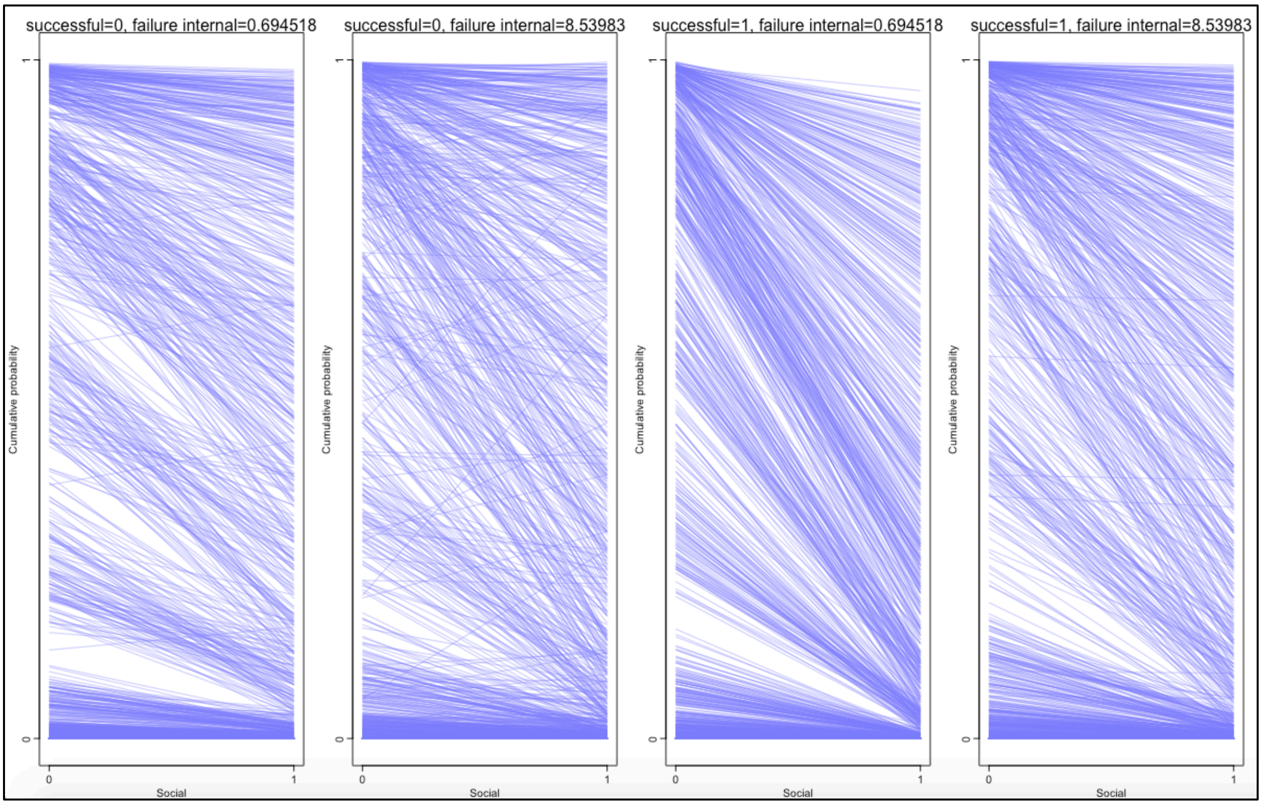

***Figure 26*. Four graphs (from left to right: A, B, C, and D) showing Model 8's predicted effects of changing an asocial model into a social model on microstructure similarity scores. The interactions between the successful (graphs C and D) versus unsuccessful models (graphs A and B) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) are the same as in Figure 25, except for participant age. In these four graphs, the participant age was instead set at one standard deviation above the mean, at 9.61 years old.**

I now consider the implications of these results for the current thesis. The second hypothesis moved from macrostructure to microstructure, exploring how microstructure similarity scores were affected when macrostructure was constrained by the close-ended task (as seen in Hypothesis 1). In doing this, I aimed to replicate findings from previous social learning studies conducted with close-ended tasks. Specifically, I predicted that the microstructure of participants' builds would be more similar to the microstructure of the model when the participants observed the relevant model (i.e., the 'social' model) rather than the irrelevant model (the 'asocial' model). The effect was predicted to be observable whether the social and asocial models were successful or unsuccessful (see Smith, Ward & Schumacher 1993; Rook 2008; Shalley & Perry-Smith 2001). Generally this prediction held true, with most data supporting Hypothesis 2.

In older children, Model 8 predicted that the social model would cause greater similarity in participants' microstructure scores regardless of whether the model was successful or unsuccessful, or whether the participants were successful or unsuccessful in building. The younger participants did show more variation, however. When the model was successful, they did show greater microstructure similarity to the model. However when the model was unsuccessful, this relationship was not present. Instead, when younger children exhibited less internal evidence of failure there appeared to be little effect of an unsuccessful social model on microstructure similarity scores at all, and when the younger children showed greater internal evidence of failure it appeared that Model 8 predicted that the unsuccessful social model would cause lower microstructure similarity scores. In other words, younger children who exhibited greater internal evidence of failure were less likely to build with a microstructure similar to the social model when they observed the unsuccessful social model. This could be interpreted as younger children being more selective of the social information they copied than the older children.

However, it is also true that (1) the positive effect of a social model on older children's microstructure similarity scores was stronger when the model was successful, and that (2) older children showed a greater positive relationship between the successful social model and microstructure similarity than the younger children. Therefore the explanation of the effect may partly be that younger children simply copied the model less, so that when they too reduced their reliance on the unsuccessful model, this was expressed as no influence of the model on microstructure similarity. This would not explain, however, the negative relationship between the presence of an unsuccessful social model and microstructure similarity for younger children with high internal evidence of failure.

It is noteworthy that higher internal evidence of failure was associated with greater dissimilarity between participant and model microstructure. This same trend was more weakly visible amongst the older children, too. It would seem at odds with the findings of the Hypothesis 1, and thus challenges both explanations given above: (A) that greater internal evidence of failure caused more reliance on social information, and (B) that in this experimental context greater reliance on social information caused lower internal evidence of failure. However, while Hypothesis 1 dealt with variation in macrostructure, Hypothesis 2 dealt with microstructure, so it is plausible (indeed hypothesised for Chapter 6) that the microstructure/macrostructure division accounts for differences in internal evidence of failure's effects. Data thus far, therefore, has indicated that higher internal evidence of failure was associated with both greater and lesser similarity scores, depending on whether the outcome variable was participants' macrostructure or microstructure. Moreover, Hypothesis 2 was generally supported, with reservations for younger children who observed the unsuccessful model.

5.3: Hypothesis 3

Hypothesis 3 stated that across close-ended social model conditions, the success of the model would be a positive predictor of variation in microstructure similarity scores. This is because the successful, but not unsuccessful, social model would show the microstructure design to be useful in tower building. After removing cases of missing data, and of participants building in open-ended conditions, the sample size used for Model 9 numbered 273. The account of the process of model comparison can be found in Appendix 7.3. The model resulting from this process is described as such:

(Model 9)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_S S_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{US} US_i + \beta_{UNG} UNG_i + \beta_{UNS} UNS_i + \beta_{UGS} UGS_i + \beta_{UNGS} UNGS_i$$

Model 9's main predictor variable was the success of the model (U), which interacts with three other variables: internal evidence of failure (N), participant age (G), and the sociality of the model (S).

Addition of more variables, such as participant sex, reduced the ability of the model to make predictions about the data. This reduction indicates these other variables did not add enough information to make them useful enough to overcome the risk of overfitting, so were probably not as important for understanding the effect of model success on microstructure similarity score variation. The marginal effect of model success in Model 9 was largely positive. The mean effect was 2.86 (SD=1.82; HPDI=0.89, between 0.06 and 5.80; see Figure 27). However, I predicted in Hypothesis 3 that model success would be a positive predictor of microstructure score variation in social conditions. If model success were also a positive predictor of microstructure score variation in asocial conditions, then it would appear that the observed similarity was due to factors other than copying.

To answer this question, I present graphs that go beyond the marginal effect of the main predictor variable. Figures 28 and 29 illustrate the predicted real effects of a successful versus unsuccessful model, dependent on interactions with the other variables of Model 9. Overall, Hypothesis 3 seems to have been supported more in older children than younger, though all conditions with the social model show a positive relationship between the successful model and microstructure similarity scores. The strongest predicted effect of model success on microstructure similarity appears to have been in older children with a social model and low internal evidence of failure (Figure 29's graph C). The second

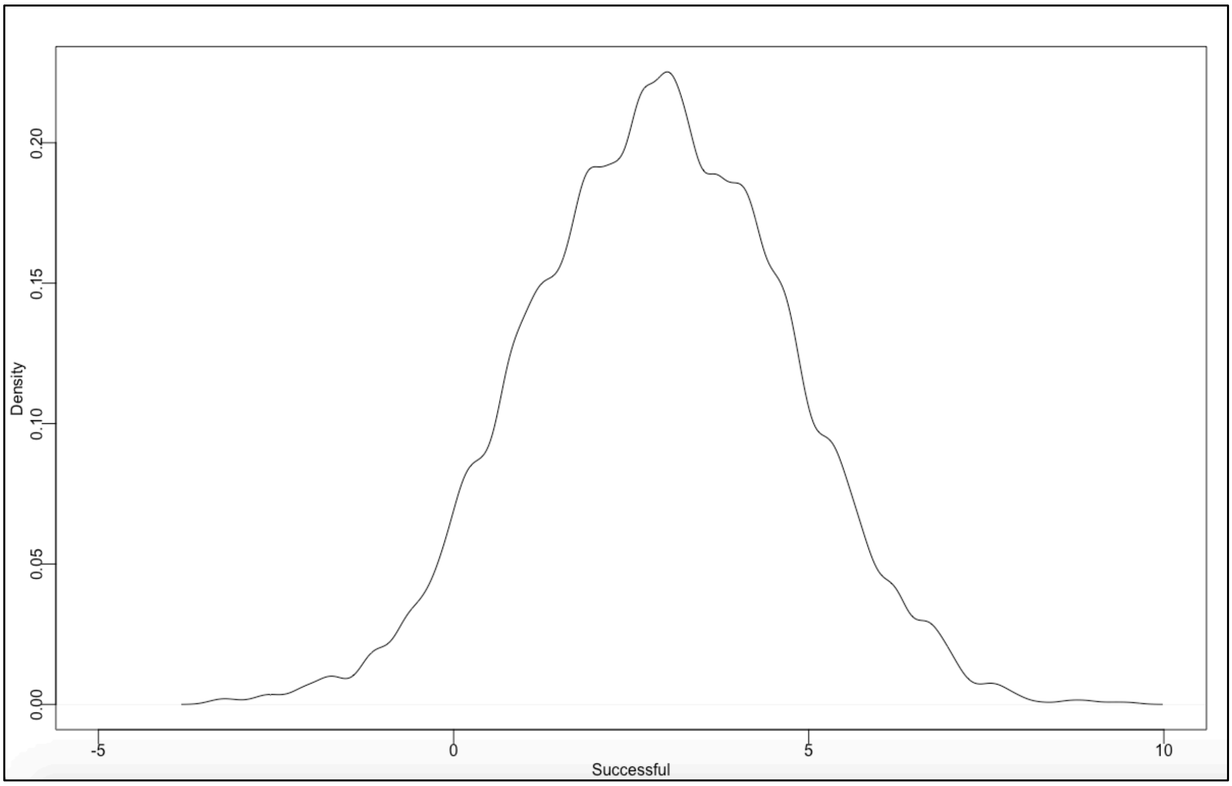

Figure 27**. Graph showing the posterior distribution of the marginal effect on microstructure similarity of the variable ('U') that indicates whether the model was successful rather than unsuccessful (in Model 9).**

strongest predicted effect of model success appears to have been younger children also with a social model and low internal evidence of failure (Figure 28's graph C). Nevertheless, younger children with low internal evidence of failure and an asocial model (graph A, Figure 28) also displayed a positive relationship between model success and microstructure similarity scores.

For older children (Figure 29) there was a clear difference between participants in the social and asocial model conditions: building with low internal evidence of failure the direction of the slope even reversed between the social and asocial models (graphs A and C respectively). The difference between social (graphs C and D) and asocial (graphs A and B) models in younger children (Figure 28) was less clear. In asocial conditions (graphs A and B), especially with low internal evidence of failure (graph A), the positive relationship between model success and microstructure similarity remained. This suggests that copying may not have been the only factor in generating the positive trend in social conditions for younger children. However it should be noted that the social model conditions (graphs C and D) do display stronger positive relationships with microstructure similarity than the equivalent asocial conditions (graphs A and B), indicating

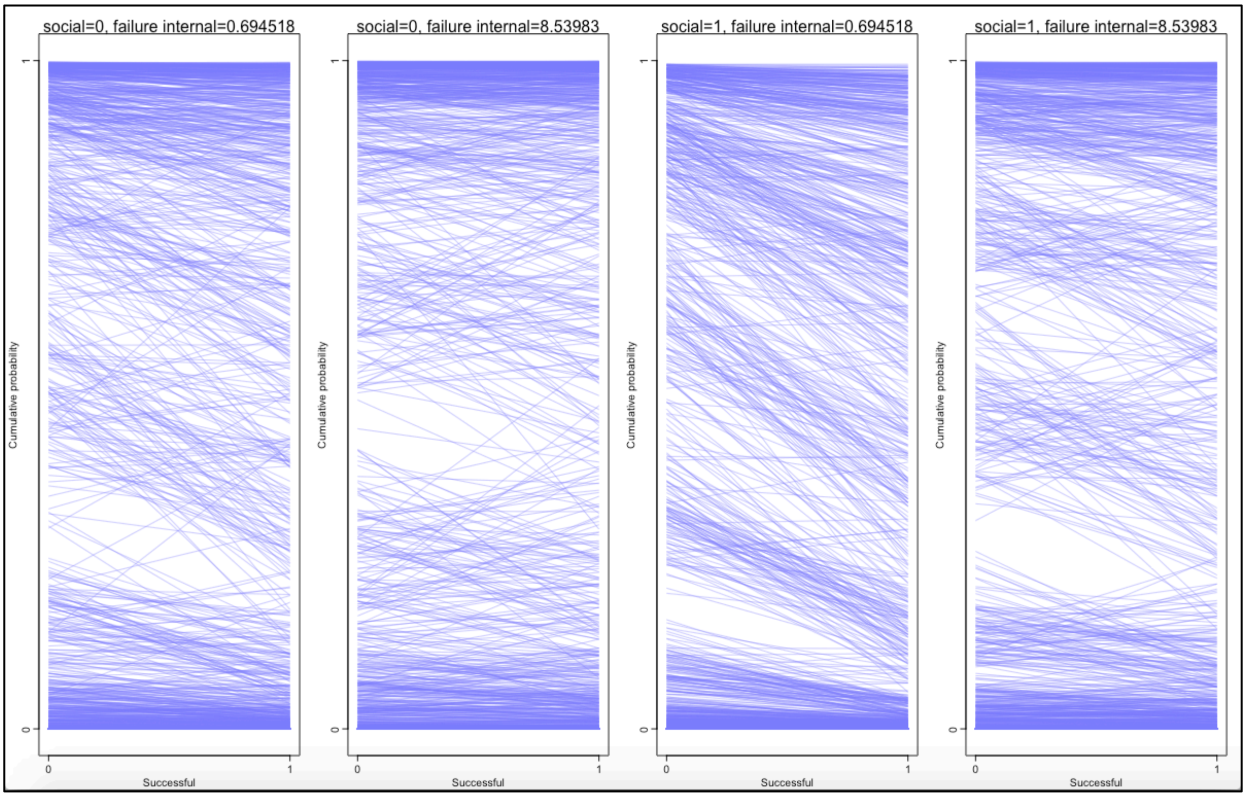

Figure 28. **Four graphs (from left to right: A, B, C, and D) showing the effect on microstructure similarity scores of changing an unsuccessful model into a successful model. The effect of model success is shown in interaction with the asocial (graphs A and B on the left) versus social (graphs C and D on the right) model, and low (graphs A and C on the far left and second from right, respectively) and high (graphs on the second from left and far right, respectively) internal evidence of failure. 'Low' internal evidence of failure was set at one standard deviation below the mean, and 'high' at one standard deviation above the mean. For all four of these graphs, participant age was set at 6.01 years old, one standard deviation below the mean age of the usable dataset.**

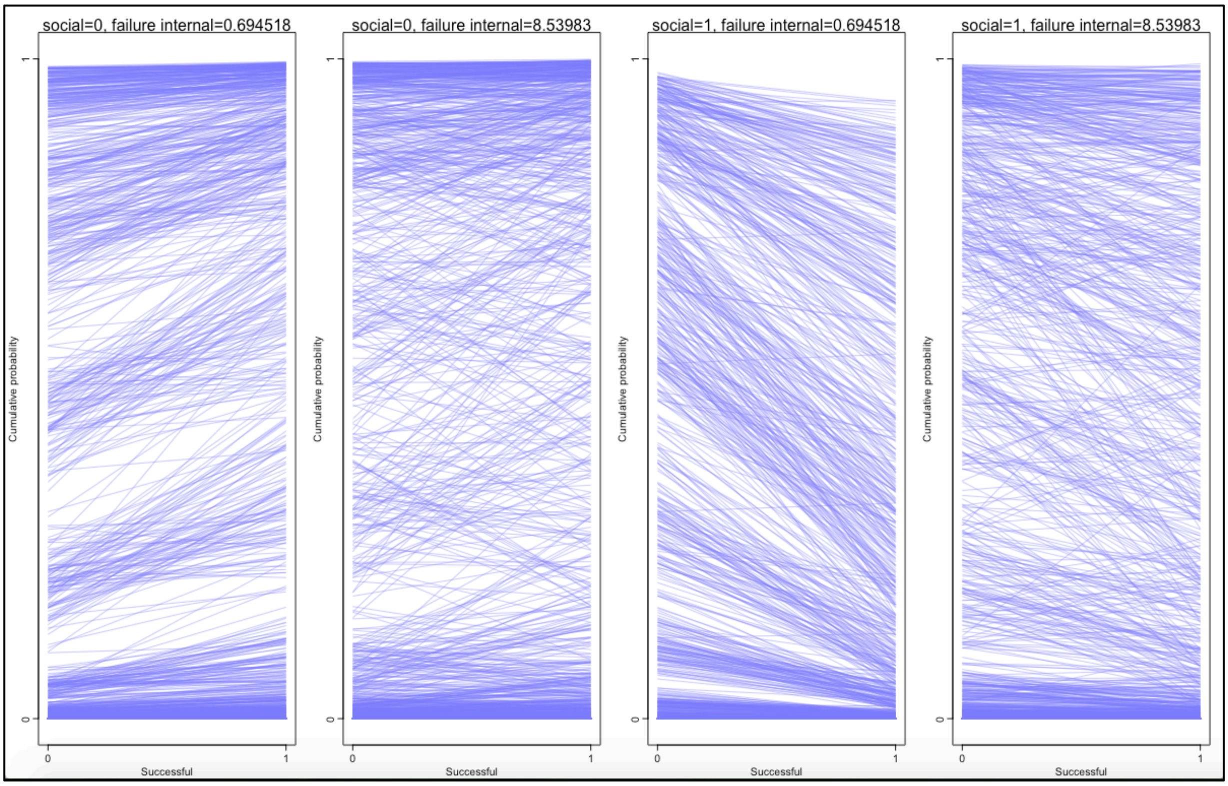

Figure 29. **Four graphs (from left to right: A, B, C, and D) showing the effect on microstructure similarity scores of changing an unsuccessful model into a successful model. The graphs illustrate the interactions between the asocial (graphs A and B) versus social model (graphs C and D) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) like Figure 28 above. However, for these four graphs, the participant's age was set at 9.61 years old, one standard deviation above the mean age of the usable dataset.**

that copying of the successful model did have some role to play for younger children. The influence of high rather than low internal evidence of failure (graphs B and D) appears to have reduced the influence of the main predictor variable on microstructure similarity. Hypothesis 3 therefore seems to have been supported by Figure 29's data for older children, whilst the effect of the successful model in Figure 28's younger children was less clear. See Appendix 8.3 for more detailed description of the results of this statistical analysis.

Hypothesis 3 therefore continued to explore variation in microstructure similarity scores under close-ended conditions by focusing on the effects of model success. Hypothesis 3, and prior literature (Kendal et al. 2005; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017), predicts the successful social model to cause increased similarity to the social model in participants' microstructure designs, and that this effect should be recognisably different from the effect of successful asocial models on participants' microstructure similarity. The positive relationships between model success and microstructure similarity across all conditions with a social model provided evidence in support of Hypothesis 3. This was bolstered in older children since, where the model was asocial, model success had negative effects on microstructure similarity. Older children thus conformed to Hypothesis 3's predictions, though high internal evidence of failure appears to have made this relationship messier than with low internal evidence of failure.

For younger children the picture was less clear due to the somewhat positive relationships between model success and microstructure similarity when the model was asocial. Nevertheless, the effect of model success was predicted to be clearer and stronger when the model was social, indicating that the positive relationship is at least partially explained by younger children's increased reliance on a social model when the model was successful rather than unsuccessful. For younger children, as for the older children, high internal evidence of failure appears to have made the positive relationships between model success and microstructure similarity less clear.

The association of high internal evidence of failure with lower microstructure similarity scores thus appears to corroborate the findings of Hypothesis 2, both of these hypotheses being concerned with variation in microstructure similarity scores. One simple explanation for the positive relationship between microstructure similarity and the successful asocial model in younger children exhibiting lower internal evidence of failure, therefore, is that internal evidence of failure was more a product of a child's building style than a cause of it, in which high internal evidence of failure was caused by both a macrostructure similar to the social model (see Hypothesis 1) and a microstructure dissimilar to the social model (Hypotheses 2 and 3). However, it does not explain why only younger children showed this positive relationship between asocial model success and microstructure similarity scores, while older children showed a negative relationship between the same variables. Overall, then, Hypothesis 3 seems to have support from the data, albeit in a clearer way from older rather than younger children, and from participants exhibiting lower rather than higher internal evidence of failure.

5.4: Hypothesis 4

Hypothesis 4 stated that across close-ended social model conditions, the success of the model would not be a good predictor of variation in macrostructure similarity scores. This is because macrostructure diversity should be constrained by the close-ended setup. The sample size of the data for Model 10, below, also numbered 273. This number was reached by subtracting the two builds for which macrostructure were not coded, the cases for which other data were absent, and those participants who built open-ended conditions (since the hypothesis did not make a prediction about these cases).

(Model 10)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_S S_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{US} US_i + \beta_{UNG} UNG_i + \beta_{UNS} UNS_i + \beta_{UGS} UGS_i + \beta_{UNGS} UNGS_i$$

In this model, variation in the macrostructure similarity score outcome variable was predicted by the main predictor, the successful (versus unsuccessful) model (U), in interaction with three other predictor variables: internal evidence of failure (N), participant age (G), and the social versus asocial model (S). As above, the addition of further predictor variables and interactions led to worse predictions (see Appendix 7.4 for greater detail). Figure 30 illustrates the centring of the mean effect of 'successful' on -0.06 (SD=2.37; HPDI=0.89, between -3.82 and 3.61). However, Model 10 includes a variable which describes the difference between social and asocial models. This variable may contribute to the wide standard deviation and HPDI, if the effect of model success varied between the social and asocial models. This can be explored further through graphing the effects of the variables on the outcome variable in interaction with one another.

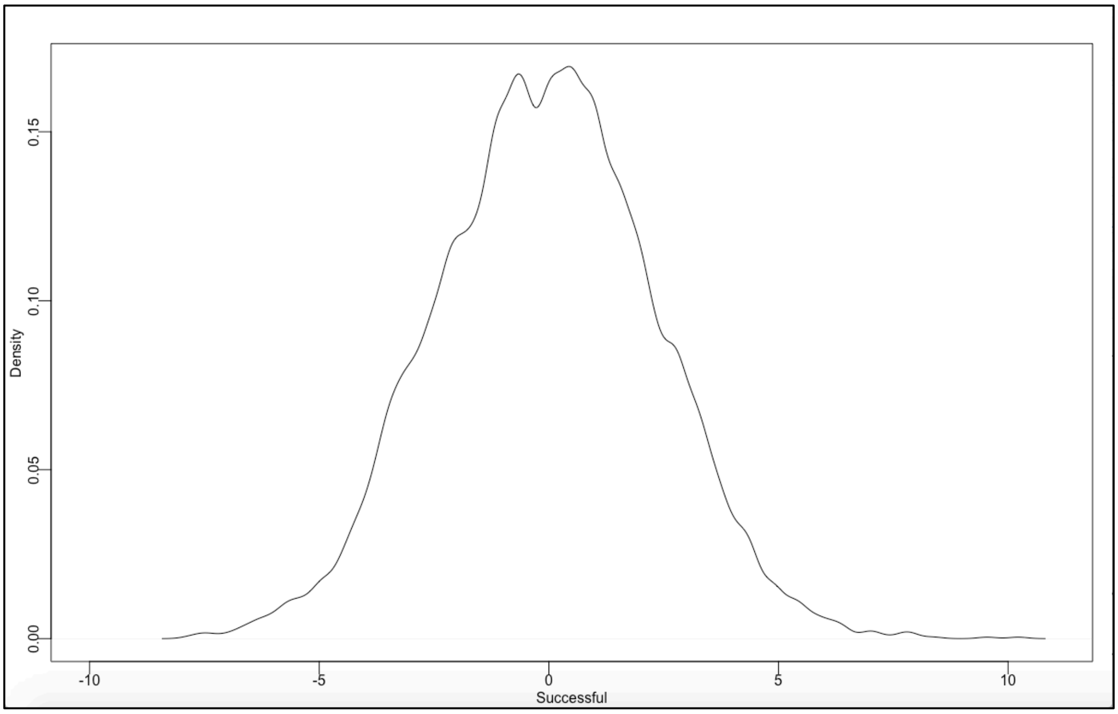

***Figure 30*. Graph showing the posterior distribution of the marginal effect of the variable for model success in Model 10.**

Figures 31 and 32 illustrate Model 10's predicted effects of changing an unsuccessful model into a successful model, in interaction with the other variables of the model: participant internal evidence of failure, participant age, and the social versus asocial model. Hypothesis 4 initially seems to have been supported in younger children (Figure 31) but challenged in older children (Figure 32). In Figure 31, the younger children's macrostructure similarity scores showed an effect of the social successful model (graphs C and D) which was no more positive than that of the asocial successful model (graphs A and B).

For younger children exhibiting high internal evidence of failure, there was a negative relationship between asocial model success and macrostructure similarity (graph B) which seems to have been made neutral when the model was social (graph D). In contrast, the older children (Figure 32) showed a clearly positive effect of a social successful model (graphs C and D), while the effect of an asocial successful model on macrostructure similarity was weaker (graphs A and B). Hypothesis 4 therefore appears to have been supported only amongst the younger children. Some further discussion of results displayed in Figures 31 and 32 can be found in Appendix 8.4.

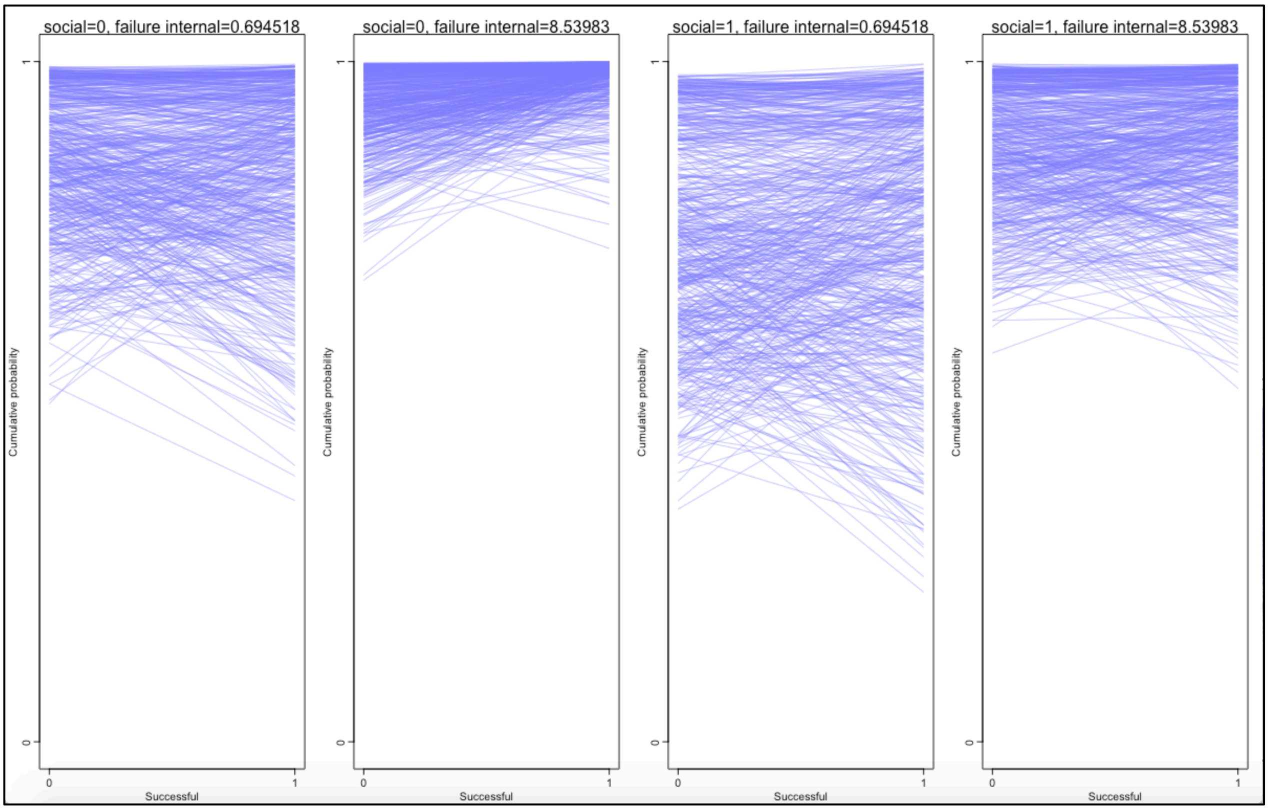

***Figure 31*. Four graphs (from left to right: A, B, C, and D) describing Model 10's predicted effects on macrostructure similarity scores of changing an unsuccessful model into a successful model. The effect of model success is shown in interaction with the asocial (graphs A and B, on the left) versus social (graphs C and D, on the right) model, and low (graphs A and C, on the far left and second from right respectively) versus high (graphs B and D, the graphs second from left and on the far right respectively) internal evidence of failure. 'Low' internal evidence of failure was set at one standard deviation below the mean, and 'high' at one standard deviation above the mean. For all four of these graphs, participant age was set at 6.01 years, one standard deviation below the mean age.**

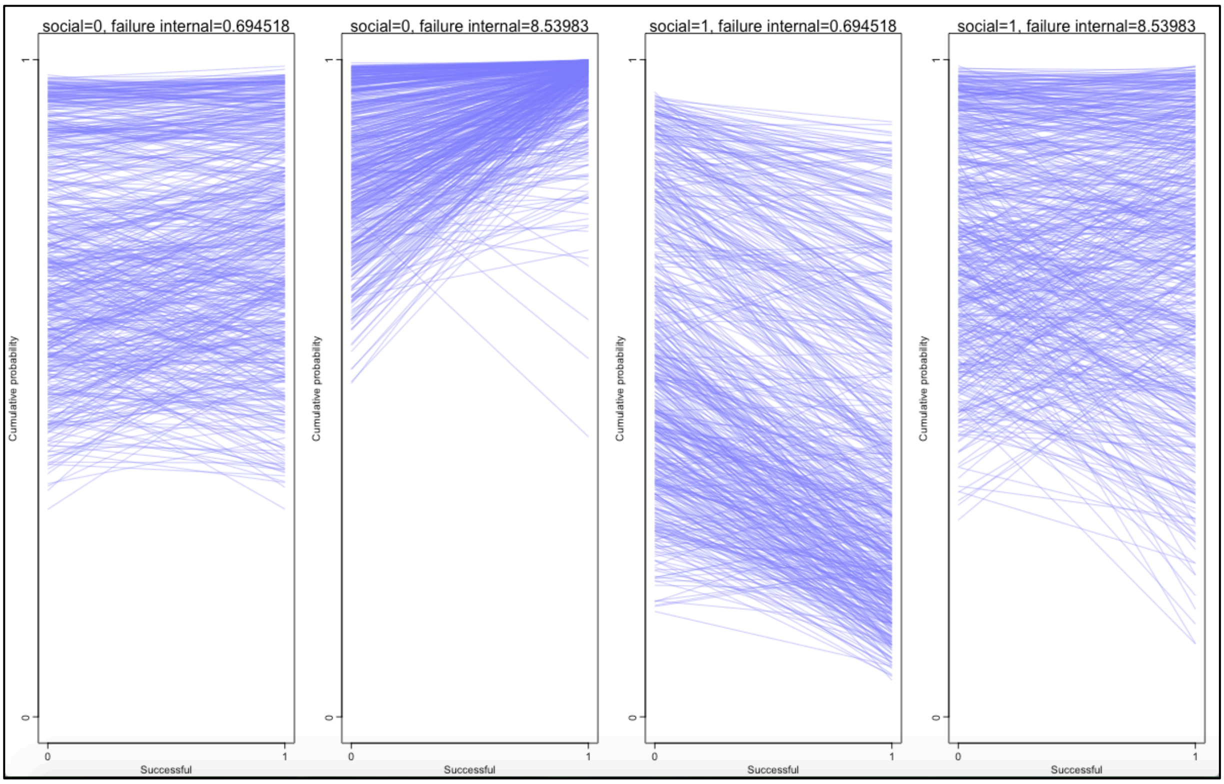

***Figure 32*. Four graphs (from left to right: A, B, C, and D) describing Model 10's predicted effect on macrostructure similarity scores of changing an unsuccessful model into a successful model. The graphs illustrate the interactions between the asocial (graphs A and B) versus social model (graphs C and D) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) like Figure 31 above. However, for these four graphs, the participant's age was set at 9.61 years old, one standard deviation above the mean age.**

The fourth hypothesis returned to macrostructure similarity scores to compare how the success of the model affected variation in this measure of structural similarity, rather than in microstructure. In Hypothesis 4 I predicted that across close-ended social model conditions, the success of the social model would not be a good predictor of variation in macrostructure similarity scores, due to the constraints of the close-ended task. If macrostructure variation were constrained by the close-ended task, there should have been little to no effect of changing an unsuccessful social model into a successful one. This prediction appeared to have been supported by data from younger but not older children.

Younger children showed no positive effect of social model success when internal evidence of failure is high, and no more positive an effect of a successful social model compared to a successful asocial model when internal evidence of failure was low. Simply, when younger children observed the successful social model they showed no increase in the macrostructure similarity of their builds to this successful social model build compared to when the model was successful yet irrelevant to their building (i.e., when the model was 'asocial'). This was not the case for older children, who did show more positive effects of a successful model when the successful model was social (i.e., relevant) rather than asocial (i.e., irrelevant). Macrostructure constraints imposed by the close-ended task thus seemed to affect younger children's macrostructure similarity more than older children's. This may be because younger children could have paid greater hindrance to the instructions given to them.

However, if this were the case, we would expect to see the blue lines on the left hand sides (i.e., with the unsuccessful rather than successful model) of each of Figure 31's graphs spread out more than the lines of Figure 32's graphs, which is not the case. Indeed when the model was unsuccessful, the older children showed greater macrostructure similarity scores than the younger children. The answer may thus lie with young children having been less influenced by social information than older children in all conditions. Hypotheses 2 and 3 both found that older children more consistently and strongly supported predictions concerning increased copying of a social/successful model than younger children. It is thus perhaps less surprising that in Hypothesis 4 younger children more readily supported a prediction of decreased similarity in macrostructure than older children.

Nevertheless, it is worth noting that the predicted positive effects of a successful social model on macrostructure similarity scores in a close-ended task (Figures 31 and 32, Hypothesis 4) were weaker than the predicted positive effects of a successful social model, also in a close-ended task, on microstructure similarity scores (Figures 28 and 29, Hypothesis 3). This suggests that, when the task was close-ended, microstructure similarity was more positively affected by changing a social model from unsuccessful to successful than macrostructure similarity scores. This gives only weak support to Hypothesis 4, with the effect having been less than what Hypothesis 4 originally predicted. The effects of model success on macrostructure similarity scores in an open-ended task, rather than the close-ended one, are explored in Chapter 6.

Chapter 6: Results and discussion for the effects of the open-ended task

In Chapter 6 I ask: 'How do children balance copying of microstructure and macrostructure within open-ended play?' I aimed to create conditions in which children combined social learning of macrostructure with so-called 'asocial' learning of microstructure, and conditions in which children combine social learning of microstructure with 'asocial' learning of macrostructure. After describing results for the four hypotheses, I interpret the results for each of the hypotheses to understand the results in terms of their implications for the current thesis. In broad terms, I argue the data indicated (1) that in the open-ended task the success of the model only increased the similarity of participants' microstructure designs when the participants themselves did not show high levels of failure in the task, (2) that in the open-ended task the effect of model success on macrostructure similarity was varied and more so than in its effect on microstructure similarity, (3) that participants demonstrating greater internal evidence of failure did not rely on greater copying of the social model's microstructure design, and (4) that participants demonstrating greater internal evidence of failure did not rely on greater copying of the social model's microstructure design. Again a predominant theme throughout these results was variation and interdependence. The hypothesised effects of the main predictor variables in each of the four hypotheses varied in strength, as well as sometimes direction, dependent on the status of other variables included in each model.

6.1: Hypothesis 1

Hypothesis 1 stated that across the open-ended social model conditions, the success of the model should positively predict variation in microstructure similarity scores, since the unsuccessful model showed the microstructure design to fail. The sample size for this hypothesis numbered 288 cases. Cases were excluded if data were missing or if participants built under close-ended conditions. As in Chapters 4 and 5, for each of Chapter 6's hypotheses a process of model comparison was completed to find the model containing interactions between the variables which best balances risks of underfitting and overfitting.

An account of the comparison process for Hypothesis 1 can be found in Appendix 9.1. The model used for this hypothesis is described as:

(Model 11)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{USN} USN_i$$

This model described variation in microstructure similarity scores through interactions between the main predictor variable, model success (U), and two other predictor variables: the social versus asocial model (S), and internal (i.e. participant) evidence of failure (N). Adding any further interactions with more predictor variables resulted in worse model predictions of future data. Figure 33 shows the posterior distribution for the marginal effect of successful models on microstructure similarity scores. The mean effect of the variable 'successful', as calculated by Model 8, was -0.22, with a standard deviation of 0.43 and 0.89 HPDI between -0.86 and 0.49, straddling the positive-negative divide.

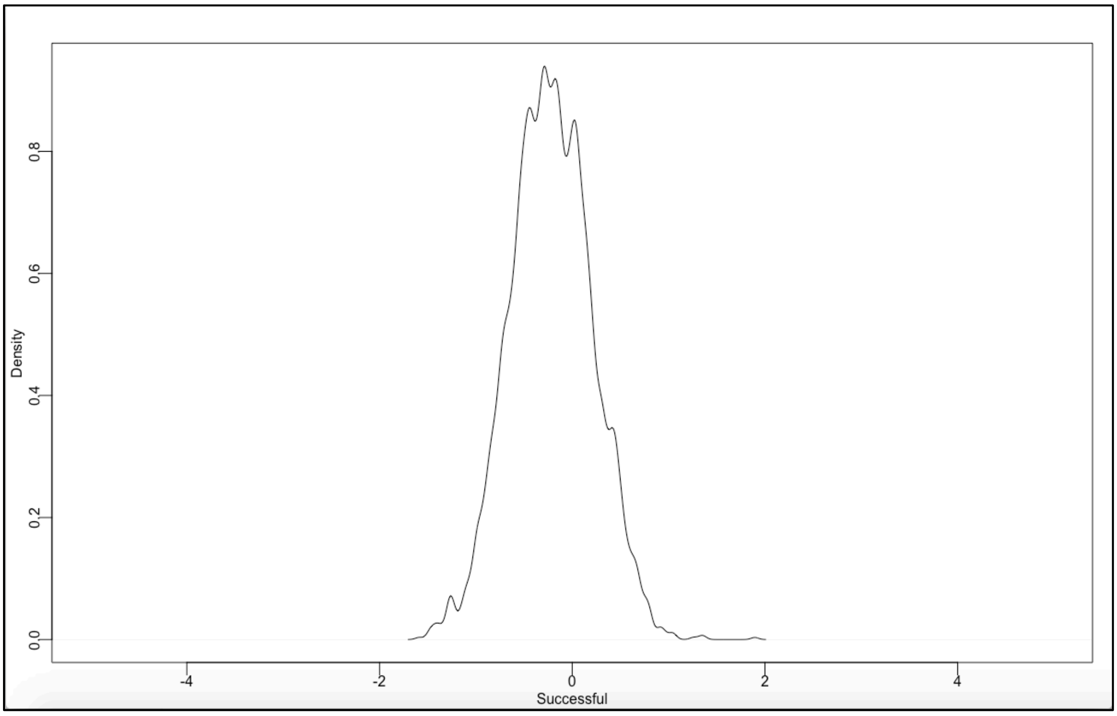

***Figure 33*. Posterior distribution of the marginal effect of the variable 'successful' in Model 11.**

Once again I used variations on the triptych plot to examine the effect of the main predictor variable on the outcome variable, in interaction with the effects of the other predictor variables. Figure 34 illustrates Model 8's predicted effects of turning an unsuccessful model into a successful model across social (graphs C and D) versus asocial models (graphs A and B), and low (graphs A and C) versus high internal evidence of failure (graphs B and D). The results were not all in line with the predictions of Hypothesis 1. There was some support for the hypothesis from results for the effect of the successful model in conditions of low internal evidence of failure (graphs A and C). In the asocial condition with low internal evidence of failure (graph A), there appears to have been a weakly negative relationship between turning an unsuccessful model into a successful one and microstructure similarity scores. This changed into a weakly positive relationship when the model was social rather than asocial (graph C). But for participants exhibiting high internal evidence of failure (graphs B and D), this comparison between asocial and social conditions did not hold. There was a similarly weak positive relationship in children exhibiting high internal evidence of failure between asocial model success and microstructure similarity scores (graph B). This was also inverted in the social model condition with high internal evidence of failure, meaning that Model 8 predicts that when there was a social model and high internal evidence of failure (graph D), model success would create less similarity in participants' builds to the model build. This is contrary to what I had predicted in Hypothesis 1.

The effects of model success on microstructure similarity appeared dependent on the effects of other variables. The influence of changing an asocial model (graphs A and B) into a social model (graphs C and D) inverted the effect of a successful model on microstructure similarity scores. The influence of changing a participant's internal evidence of failure from low (graphs A and C) to high (graphs B and D) appears to have also inverted the effect of a successful model on microstructure similarity scores. More detailed description of the effects of these interdependencies can be found in Appendix 10.1.

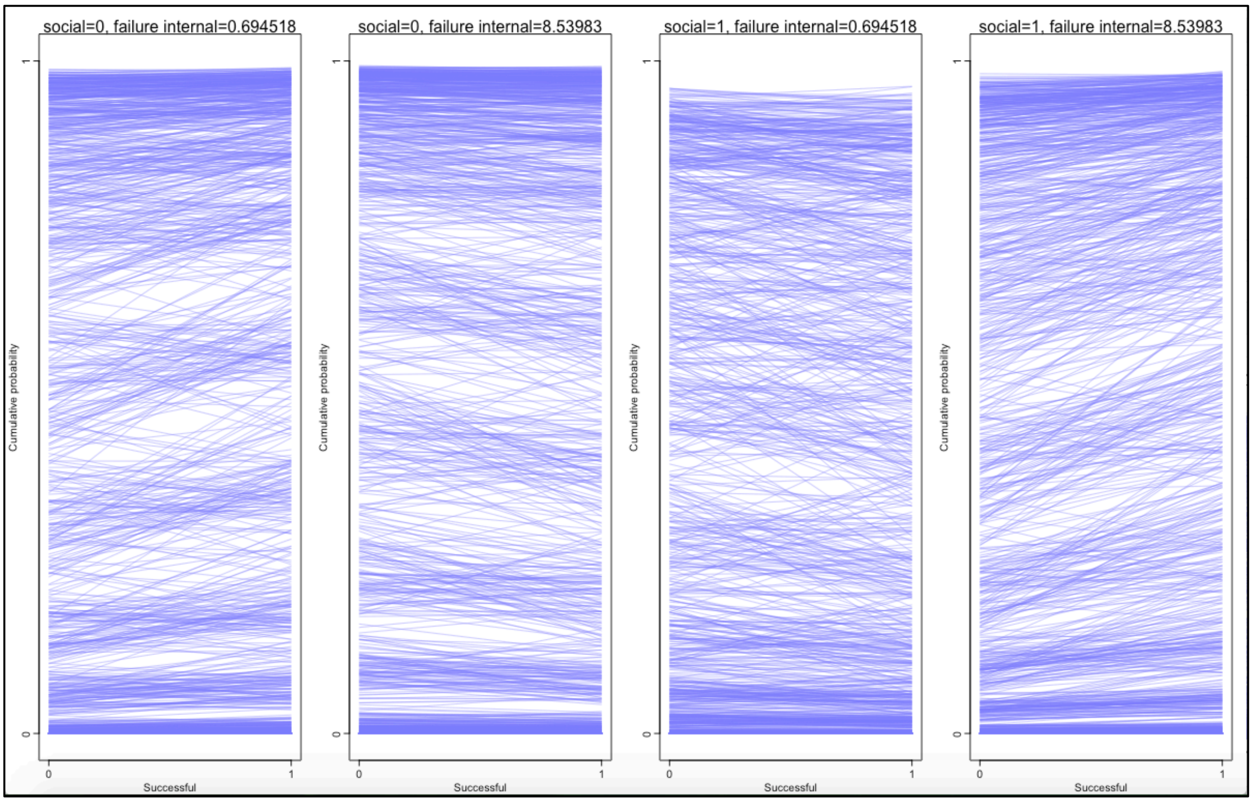

***Figure 34*. Four graphs (from left to right: A, B, C, and D) illustrating Model 11's predictions for the effect on microstructure similarity of turning an unsuccessful model into a successful model. Graphs A and B (on the left) show the effect of 'successful' when the model was asocial, while graphs C and D (on the right) show the effect of 'successful' when the model was social. Graphs A and C (on the far left and second from right) show the effect of 'successful' when the participant exhibited low internal evidence of failure, and graphs B and D (on the second from left and far right) show the effect of 'successful' when the participant exhibited high internal evidence of failure. 'Low' internal evidence of failure was set one standard deviation below the mean evidence of failure for the entire usable dataset, while 'high' was set one standard deviation above this mean evidence of failure.**

In Hypothesis 1 I predicted that across open-ended conditions with a social model, the success of the model should positively predict variation in microstructure similarity scores, since the unsuccessful model showed the microstructure design to fail. This was therefore the same hypothesis as Hypothesis 3 of Chapter 5, except here the children participated in an open-ended task rather than a close-ended task. Nevertheless, in this open-ended context similar positive relationships were expected between social model success and microstructure similarity as were found in Chapter 5. Since participants building with an asocial model were included also, changes between the successful and unsuccessful social models could be compared between the equivalent asocial models to assess the significance of any effect.

The hypothesis was supported only when participants exhibited low levels of internal evidence of failure. When participants did show low internal evidence of failure, asocial model success was predicted to cause children to build structures with slightly less microstructure similarity to the model they did not observe (i.e., the social model). When the model was social, however, participants with low internal evidence of failure demonstrated slightly greater microstructure similarity in their builds to the social model when that model was successful rather than unsuccessful. This is the relationship, albeit in a very weak form, predicted in Hypothesis 1. It therefore corroborates the findings of Chapter 5's Hypothesis 3 in the close-ended task, and is in line with previous literature indicating (1) that copying should be beneficial to individuals, and sustainable at a population-level, only when a learner can copy information which is useful (Giraldeau, Valone & Templeton 2002; Kendal et al. 2005; Truskanov and Prat 2018; Enquist, Eriksson & Ghirlanda 2007) and (2) that children are able to alter their reliance on social information to copy more when the model is reliable rather than unreliable, and efficient rather than inefficient (Pinkham & Jaswal 2011; Birch, Vautheir & Bloom 2008; Bandura 1986; Clement, Koenig & Harris 2004; Ma & Ganea 2010; Carr, Kendal & Flynn 2015; Turner, Giraldeau & Flynn 2017).

However, this finding was not the case for participants that exhibited high internal evidence of failure. In fact, participants demonstrating high internal evidence of failure showed the reverse effects of model success across both asocial and social models. Model 11 (see Figure 34) predicted that the success of the asocial model causes participants exhibiting high internal evidence of failure to show increased microstructure similarity to the model which these participants did not observe. This was not the same finding as reported in Hypotheses 2 and 3 of Chapter 5, which found higher internal evidence of failure associated with lower microstructure similarity scores in the close-ended task. This new data, from the open-ended task, thus challenges the simple explanation given in Chapter 5, where internal evidence of failure was considered the product of a microstructure which was dissimilar to that demonstrated by the social model. Either this explanation is merely unhelpful, or it is in some way inverted by the change from a close-ended to an open-ended task.

This inversion idea is itself challenged by the finding that when the model was social and participant evidence of failure was high, model success was negatively related to microstructure similarity scores. This result does fit the internal evidence of failure explanation from Chapter 5, that internal evidence of failure was made more likely by a microstructure which less resembled that used by the social model. This result also contradicts the literature cited above, since it suggests that participants who saw their own building to be more ineffective than others relied less on social information. It can be argued that this could make sense if the internal evidence of failure score were treated only as an outcome of how similar a child's building was to the model's: that participants who attempted to do something different from the model encountered greater internal evidence of failure than participants who copied the model. However, existing literature informs us that children theoretically should (Feldman, Aoki & Kumm 1996) and actually do (Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a) rely on copying a social model more when their own solutions are shown to be ineffective in a given task. Why this should not be the case here is currently unclear. The role of internal evidence of failure is investigated further in Hypotheses 3 and 4 below. Overall it appears that in the open-ended task, when participants encountered low internal evidence of failure, model success increased microstructure similarity scores only weakly, and that when participants demonstrated higher levels of internal evidence of failure model success reduced microstructure similarity scores.

6.2: Hypothesis 2

In Hypothesis 2 I predicted that across open-ended social model conditions, model success would not be a positive predictor of macrostructure similarity scores. For this hypothesis, the dataset numbered 286 cases: excluding those with missing values and those collected from the close-ended task. Model 12, described below, was the product of the same process of model comparison described above. A fuller description of this process for Hypothesis 2 can be found in Appendix 9.2.

(Model 12)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_T T_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UT} UT_i + \beta_{USN} USN_i + \beta_{USG} USG_i + \beta_{UST} UST_i + \beta_{UNG} UNG_i + \beta_{UNT} UNT_i + \beta_{UGT} UGT_i + \beta_{USN} USN_i + \beta_{USNG} USNG_i + \beta_{USNT} USNT_i + \beta_{USGT} USGT_i + \beta_{UNGT} UNGT_i + \beta_{USNGT} USNGT_i$$

Model 9 describes interactions between the main predictor variable, a successful rather than unsuccessful model (U), and three other predictor variables: the social rather than asocial model (S), internal evidence of failure (N), the age of the participant (G), and the degree of the participant's attendance to the video (T). Model 12's predicted marginal effect of model success on macrostructure similarity scores had a quite strongly positive mean effect of 3.83, though with a standard deviation of 3.64 and wide 0.89 HPDI between -2.16 and 9.38. Figure 35 visualises its wide marginal distribution over both sides of zero.

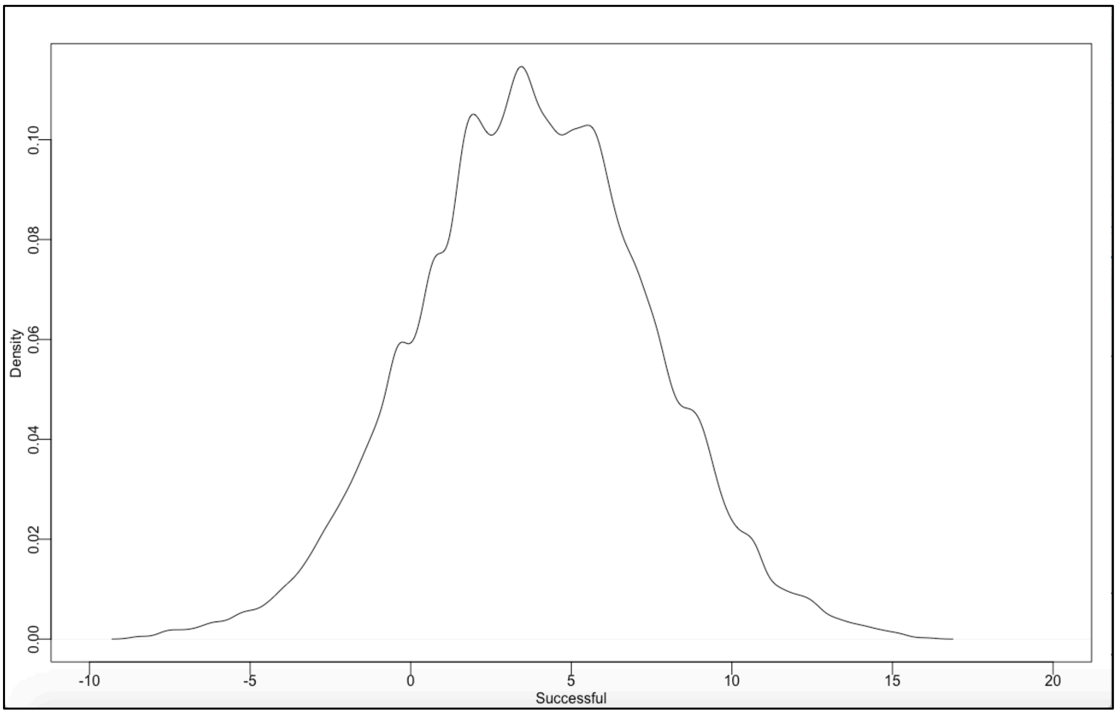

***Figure 35*. Graph showing Model 12's posterior distribution for the marginal effect of the variable 'successful'.**

Figures 36, 37, 38, and 39 illustrate Model 12's predicted effects on macrostructure similarity scores of turning the unsuccessful model into the successful model. Within each of the graphs, the effect of 'successful' was mediated by conditions of either a social or asocial model, and low or high internal evidence of failure. Figures 36 and 38 show the effect of these interactions when the participants were younger, while Figures 37 and 39 show the effect of these interactions when the participants were older. Figures 36 and 37 show the effect of all of these interactions when the participants exhibited low attendance to the experimental video, while Figures 38 and 39 show the effect of all of these interactions when the participants exhibited high attendance to the experimental video.

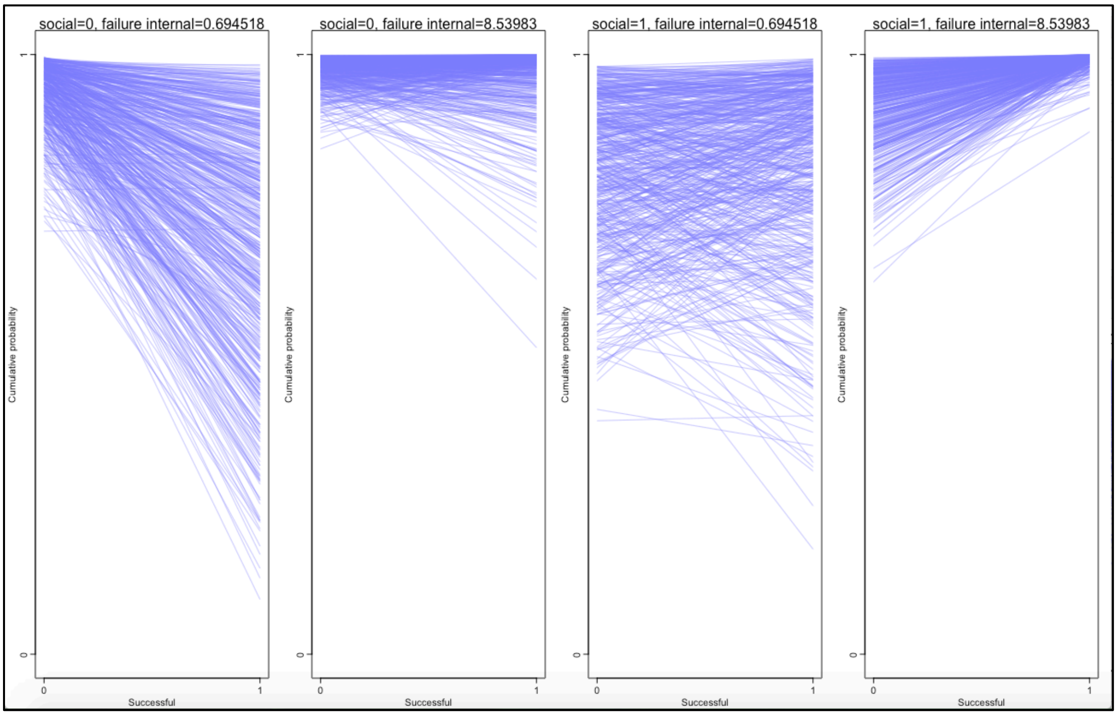

***Figure 36*. Four graphs (from left to right: A, B, C, and D) showing Model 12's predicted effects on macrostructure similarity scores of turning an unsuccessful model into a successful model. Graphs A and B (on the left) show the effect of 'successful' when the model was asocial, while graphs C and D (on the right) show the effect of 'successful' when the model was social. Graphs A and C (on the far left and second from right) show the effect of 'successful' when the participant exhibited low internal evidence of failure, and graphs B and D (on the second from left and far right) show the effect of 'successful' when the participant exhibited high internal evidence of failure. 'Low' internal evidence of failure was set one standard deviation below the mean evidence of failure of the entire usable dataset, while 'high' was set one standard deviation above the mean evidence of failure. For this set of graphs, the participant age was set at 6.01, one standard deviation below the mean participant age of the entire usable dataset. Also, for these graphs, participant attendance to the video was set at 2.44, one standard deviation below the mean score.**

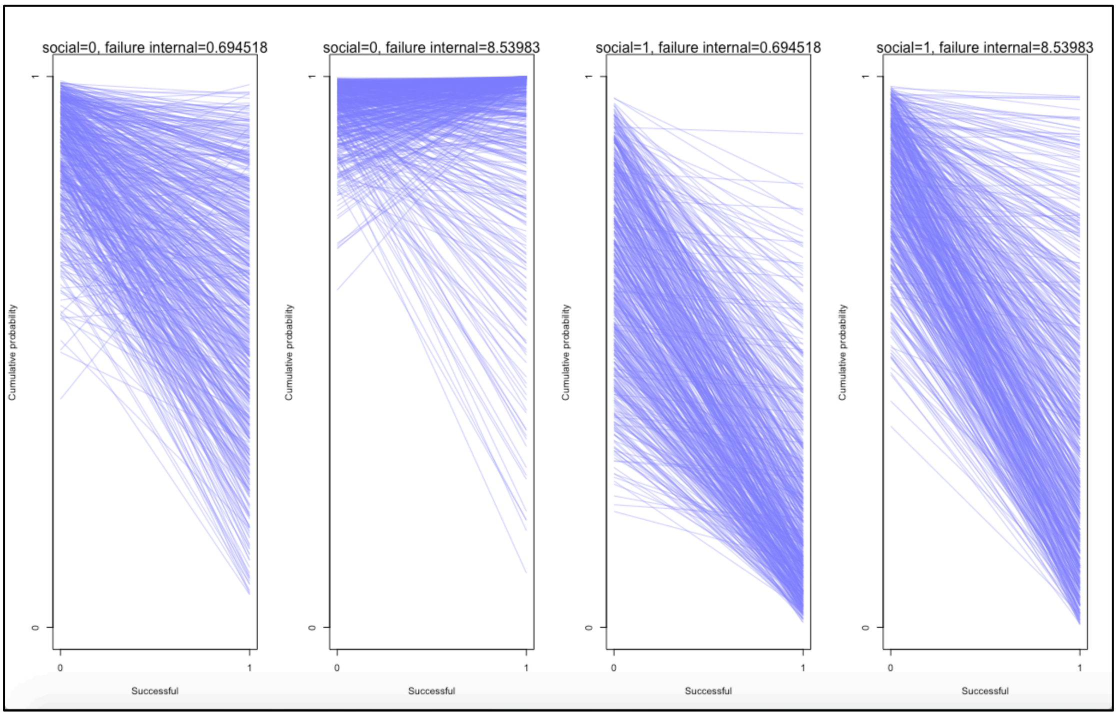

***Figure 37*. Four graphs (from left to right: A, B, C, and D) showing Model 12's predicted effects on macrostructure similarity scores of turning an unsuccessful model into a successful model. The interactions between social (graphs C and D) versus asocial models (graphs A and B) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) are the same as Figure 36 above. However for this set of graphs, while participant attendance to the video was again set to 'low', participant age was set to 9.61, one standard deviation above the mean age of the entire usable dataset.**

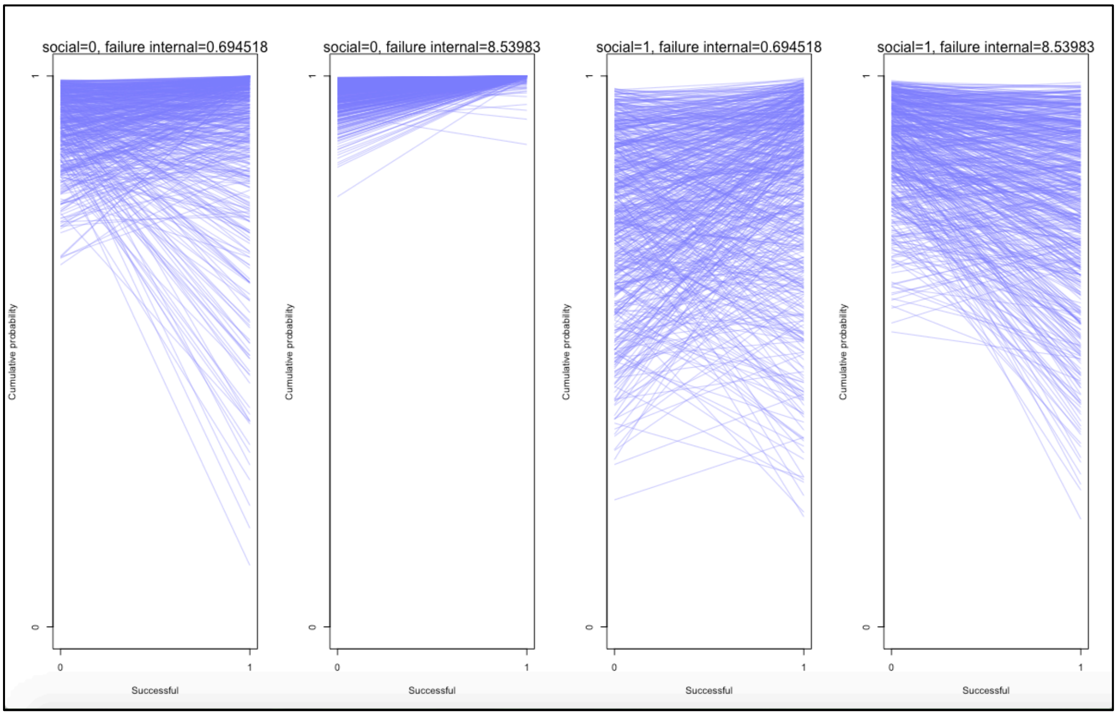

***Figure 38*. Four graphs (from left to right: A, B, C, and D) showing Model 12's predicted effects on macrostructure similarity scores of turning an unsuccessful model into a successful model. The interactions between social (graphs C and D) versus asocial models (graphs A and B) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) are the same as Figure 36 above. However for this set of graphs, while participant age was set to 'low', participant attendance to the video was set to 11.53, one standard deviation above the mean score for the entire usable dataset.**

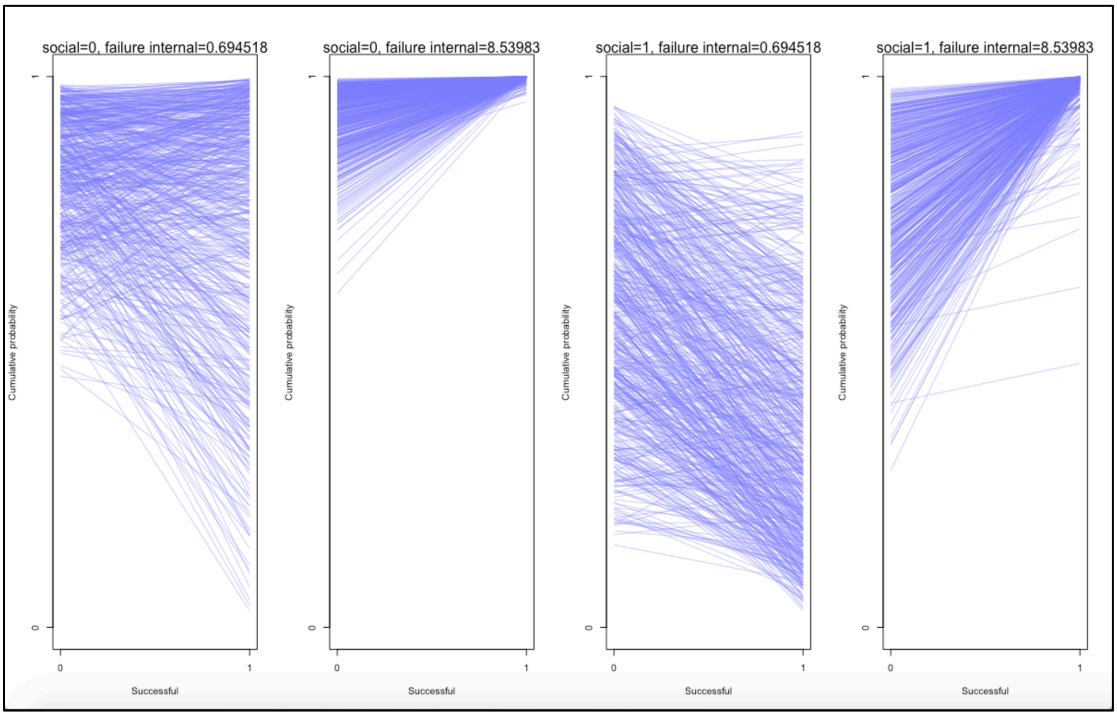

***Figure 39*. Four graphs showing Model 12's predicted effects on macrostructure similarity scores of turning an unsuccessful model into a successful model. The interactions between social (graphs C and D) versus asocial models (graphs A and B) and low (graphs A and C) versus high internal evidence of failure (graphs B and D) are the same as Figure 36 above. However for this set of graphs, both participant age and participant attendance to the video were set to 'high': one standard deviation above their mean scores for the entire usable dataset.**

Two out of the eight conditions with a social model appeared to uphold Hypothesis 2, where the transformation of an unsuccessful model into a successful model resulted in no real change in macrostructure similarity scores: in Figure 36's graph C with younger children exhibiting low attendance to the video and low internal evidence of failure, and in Figure 38's graph C with younger children exhibiting high attendance to the video and low internal evidence of failure. Furthermore, there were another two conditions, with a social model, where there was a negative effect on macrostructure similarity of turning an unsuccessful model into a social model: in Figure 36's graph D with younger children exhibiting low attendance to the video and high internal evidence of failure, and in Figure 39's graph D with older children exhibiting high attendance to the video and also high internal evidence of failure. There were then four conditions, with a social model, where, contrary to Hypothesis 2, the transformation of an unsuccessful social model into a successful social model results in increased macrostructure similarity scores: in Figure 37's graph C and D with older children exhibiting low attendance to the video and both low and high internal evidence of failure, in Figure 38's graph D with younger children exhibiting high attendance to the video with high internal evidence of failure, and in Figure 39's graph C with older children exhibiting high attendance to the video and low internal evidence of failure. Results thus give a mixed picture of the effect of a successful model on macrostructure similarity scores. When in interaction with other variables, model success did not reliably increase the macrostructure similarity of participants' builds, and in Figure 36, with young children exhibiting low attendance to the video and high internal evidence of failure (i.e., between graphs B and D), the successful social model reverses the apparently positive effect of a successful asocial model on macrostructure similarity scores.

The influence of turning an asocial model into a social model on the effect of model success on macrostructure similarity was also mixed. There were four out of eight conditions in which the social model seems to have strengthened the effect on macrostructure of whatever the effect of the successful asocial model already was: in Figure 37's graphs C and D with older children exhibiting low attendance to the video and both low and high internal evidence of failure, and in Figure 39's graphs C and D with older children exhibiting high attendance to the video and both low and high internal evidence of failure. There were then also two conditions in which the change from the asocial to social model reversed the effect of a successful model on macrostructure similarity scores: between Figure 36's graphs B and D with young children exhibiting low attendance to the video and high internal evidence of failure, and between Figure 38's graphs B and D with young children exhibiting high attendance to the video and also high internal evidence of failure. The two remaining conditions showed the influence of a social rather than an asocial model to be to make a relatively strong effect of a successful model on macrostructure scores in the asocial condition become weaker. This is true between Figure 36's graphs A and C with younger children exhibiting low attendance to the video and low internal evidence of failure, and between Figure 38's graphs A and C with younger children exhibiting high attendance to the video and low internal evidence of failure.

The influence of high internal evidence of failure, rather than low internal evidence of failure, on the effect of changing an unsuccessful model into a successful model appears to have been various, across social versus asocial models, lower and higher internal evidence of failure, and younger and older children. The influence of participant age on the effect of model success on macrostructure also appears to have been variable, and, while the effect of internal evidence of failure appears to have been quite uniform with the asocial model, this variability returned when the model was social. Overall, therefore, the picture of the effect of social model success on macrostructure similarity scores is difficult. There appear to have been many important factors, the independent effects of which are difficult to understand. A more detailed account of the complexities of the interactions between the variables can be found in Appendix 10.2.

The second hypothesis used all of the same variables as Hypothesis 1, except the outcome variable was changed from variation in microstructure to variation in macrostructure. In Hypothesis 2 I predicted that across open-ended social model conditions, model success would not be a positive predictor of macrostructure similarity scores. Out of eight conditions with a social model graphed between Figures 36, 37, 38, and 39, four did not display a positive relationship between model success and macrostructure similarity scores. However Figure 39's graph D, with older children exhibiting higher attendance to the video, shared its negative relationship between model success and macrostructure similarity with Figure 39's asocial model too (graph B), though with the social model it was stronger. The presence of negative relationships between model success and macrostructure similarity scores in the open-ended task was different from the relationships observed between model success and microstructure similarity scores in Chapter 5's close-ended task (Hypothesis 3). Furthermore, of the four conditions in which a positive relationship existed between social model success and macrostructure similarity scores, three shared this positive relationship with the equivalent asocial model conditions.

This may be considered evidence in support of a difference between how children copy macrostructural information in close-ended tasks whilst not copying macrostructure in an open-ended task. However, the aim of Chapter 6's Hypotheses 1 and 2 was to create conditions in which the same children, in an open-ended (i.e. more playful; Bateson 2014; Bateson & Martin 2013) task, copied the successful model's microstructure and the unsuccessful model's macrostructure. Results in line with this prediction would support the argument that children do copy information from models demonstrating failure (Meltzoff 1995; Huang & Charman 2005; Carr, Kendal & Flynn 2015). But this is confounded by the lack of strongly positive relationships found in Chapter 6's Hypothesis 1, above, between model success and microstructure similarity. Indeed there was no difference between the negative effects of model success in older children demonstrating high internal evidence of failure on microstructure and macrostructure similarity. The effects predicted by Hypothesis 2 were therefore only present under specific conditions.

In fact, the overriding impression from Hypothesis 2's results is the large degree of interdependence between all of the variables. It appears that children's building, under these conditions, was influenced by a large number of factors which determined whether the difference between a successful and unsuccessful model (social *or* asocial) had a positive effect, negative effect, or neutral effect on how similar their macrostructure design was to a model they may or may not have seen. This variation was even greater among conditions with the social model than conditions with the asocial model. For example, the strongly positive relationship between model success and macrostructure similarity scores in Figure 37's graph A could be turned into a strongly negative relationship by changing participant age from high to low, could be turned into a messy relationship with both positive and negative signals by making the model asocial rather than social, could be turned into an even stronger positive relationship by turning internal evidence of failure from high to low, and could be turned into another strongly negative relationship by changing attendance to the experimental video from low to high. In fact the need to have such a number of graphs illustrating predicted interactions between these variables lies in

Appendix 9.2's model comparisons, which showed that the predictions of the model could be improved by adding further variables to a point beyond which addition of variables to the other models of Chapters 5 and 6 showed similar improvements. None of the other models in Chapters 4, 5, or 6 had the number of variables as did Model 12. It therefore appears that the effects of model success on macrostructure similarity scores were more dependent on the effects of other variables than for microstructure similarity scores in Hypothesis 1's open-ended task, or for either macrostructure or microstructure similarity scores in Chapter 5's close-ended task.

6.3: Hypothesis 3

Hypothesis 3 stated that across open-ended conditions, the rate of children's internal evidence of failure would be a positive predictor of microstructure similarity scores. Like the first hypothesis of Chapter 6, which also dealt with microstructure similarity, the sample size here numbered 288 cases.

Model 13, below, described the interaction of the main predictor variable, internal evidence of failure (N), with one other predictor variable: the social versus asocial model (S). This model is the product of the process of model comparison which can be found in Appendix 9.3. Like for the other models above, addition of further parameters weakened the ability of the model to make predictions.

(Model 13)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_{NS} N S_i$$

Samples from the posterior distribution of the parameter for the marginal effect of 'internal evidence of failure' are displayed in Figure 40. Notice the very different scale of the graph compared to Figure 35 in Hypothesis 2. This posterior distribution was much more compact, with a higher maximum posterior density. The mean effect of the 'internal evidence of failure' parameter was -0.03, with a standard deviation of 0.03, and 0.89 HPDI from -0.08 to 0.03.

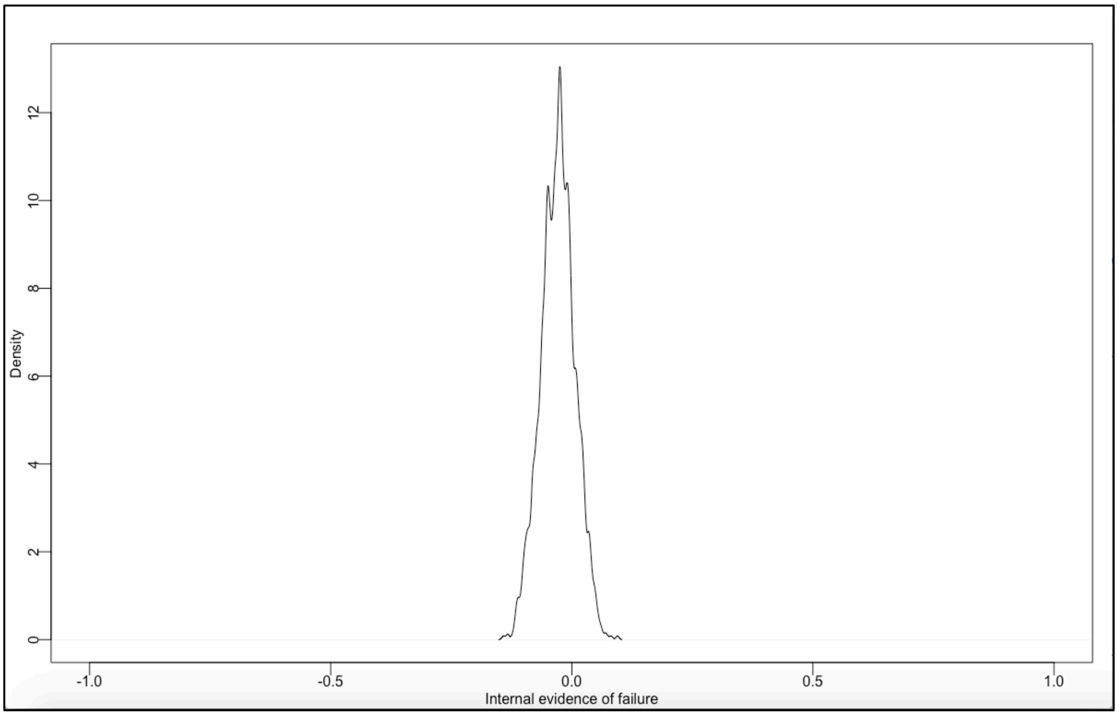

***Figure 40*. Graph displaying samples from the posterior distribution of the parameter for the marginal effect of the variable 'internal evidence of failure' in Model 13.**

Figure 41 shows Model 13's predicted effects on microstructure similarity scores of changing a participant's internal evidence of failure from one standard deviation below the mean internal evidence of failure score to one standard deviation above the mean score. The two graphs show the effect of this change on microstructure across two conditions: with either an asocial or a social model.

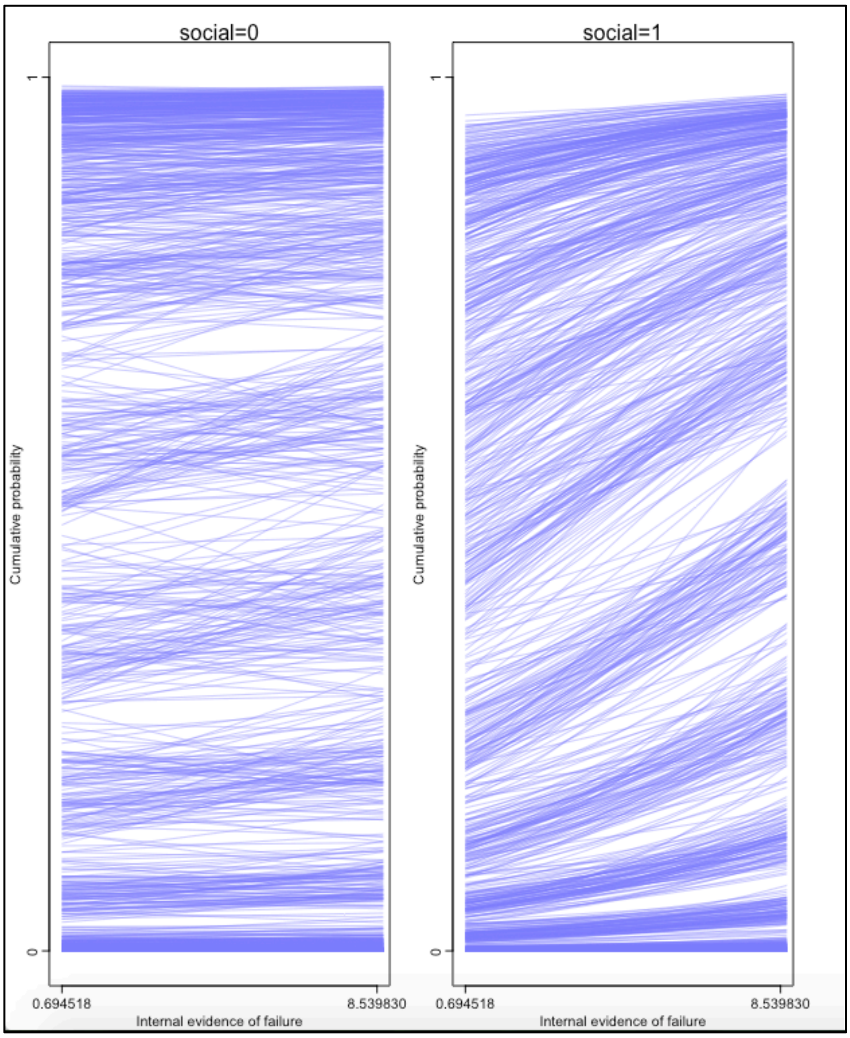

***Figure 41*. Two graphs (from left to right: A and B) illustrating Model 13's predicted effects on microstructure similarity scores of increasing participant evidence of failure from 'low' to 'high'. 'Low' internal evidence of failure was set at one standard deviation below the mean internal evidence of failure for the entire usable dataset (4.62), while 'high' internal evidence of failure was set at one standard deviation above this mean. Graph A (on the left) shows the effect of this change when the model was asocial, while graph B (on the right) shows its effect when the model was social.**

Graph B (on the right of Figure 41), where the model was social instead of asocial, does not support Hypothesis 3. This is because the effect of transforming low internal evidence of failure into high internal evidence of failure was predicted to decrease microstructure similarity scores. This was further supported by Figure 41's graph A, which predicts the effect of changing low to high internal evidence of failure, when the model was asocial, to be much weaker. There does appear to be some slight negative relationship between higher internal evidence of failure and higher microstructure similarity scores in graph A, but this is so weak an effect as to be nearly insignificant.

Hypothesis 3 moved from model success (i.e., external evidence of failure) as the primary predictor of variation in the outcome variable to the success of the participants (i.e., internal evidence of failure). Prior literature indicates (at least in close-ended experiments) that high internal evidence of failure should cause children to defer to a model (e.g., Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a; Caldwell & Millen 2010). As prior literature uses close-ended tasks, I predicted in Hypothesis 3 that across open-ended conditions, the rate of children's internal evidence of failure would be a positive predictor of microstructure similarity scores. The role of internal evidence of failure in the results so far, through Chapters 5 and 6 with both close- and open-ended tasks, has appeared complex. Yet the lack of other variables for this hypothesis in Model 13 indicates that the effect of transforming low internal evidence of failure into high was relatively invariable across conditions. When the model was asocial, the effect of internal evidence of failure on microstructure similarity was weakly negative. When the model was social, the effect of increased internal evidence of failure was more strongly negative.

This was contrary to what I predicted in Hypothesis 3 and to the prior literature on which Hypothesis 3 was based (e.g., Feldman, Aoki & Kumm 1996; Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a). However, it supports the interpretations of the role of the 'internal evidence of failure' variable in Chapters 5 and 6: that high internal evidence of failure was associated with lower microstructre similarity scores. If the effect of internal evidence of failure on microstructure similarity were neutral, then the argument could be made that children had no incentive to copy from others in the open-ended task. However, the effect of internal evidence of failure became more strongly negative with a social model. This suggests that rather than higher internal evidence of failure being exhibited merely due to microstructure designs which were more dissimilar to the social model's, that instead participants who were exposed to the social model's microstructure and who yet built with a different microstructure encountered greater internal evidence of failure.

6.4: Hypothesis 4

In Hypothesis 4 I predicted that, in open-ended conditions, the rate of children's internal evidence of failure would not be a positive predictor of macrostructure similarity scores. The sample size for this hypothesis numbered 286 cases, after excluding those participants who built under close-ended conditions and those for whom data were missing. Model 14, described below, was once again the product of a model comparison process and this is summarised in Appendix 9.4.

(Model 14)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_T T_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NT} NT_i + \beta_{NSG} NSG_i + \beta_{NST} NST_i + \beta_{NGT} NGT_i + \beta_{NSGT} NSGT_i$$

Model 14 predicts variation in macrostructure similarity scores by computing interactions between four variables: the main predictor variable, 'internal evidence of failure' (N), and the other predictor variables of 'social' (S), 'age' (G), and 'attendance to the video' (T). Figure 42 describes samples from the posterior distribution of the marginal effect of 'internal evidence of failure' on macrostructure similarity scores. The mean effect of this 'internal evidence of

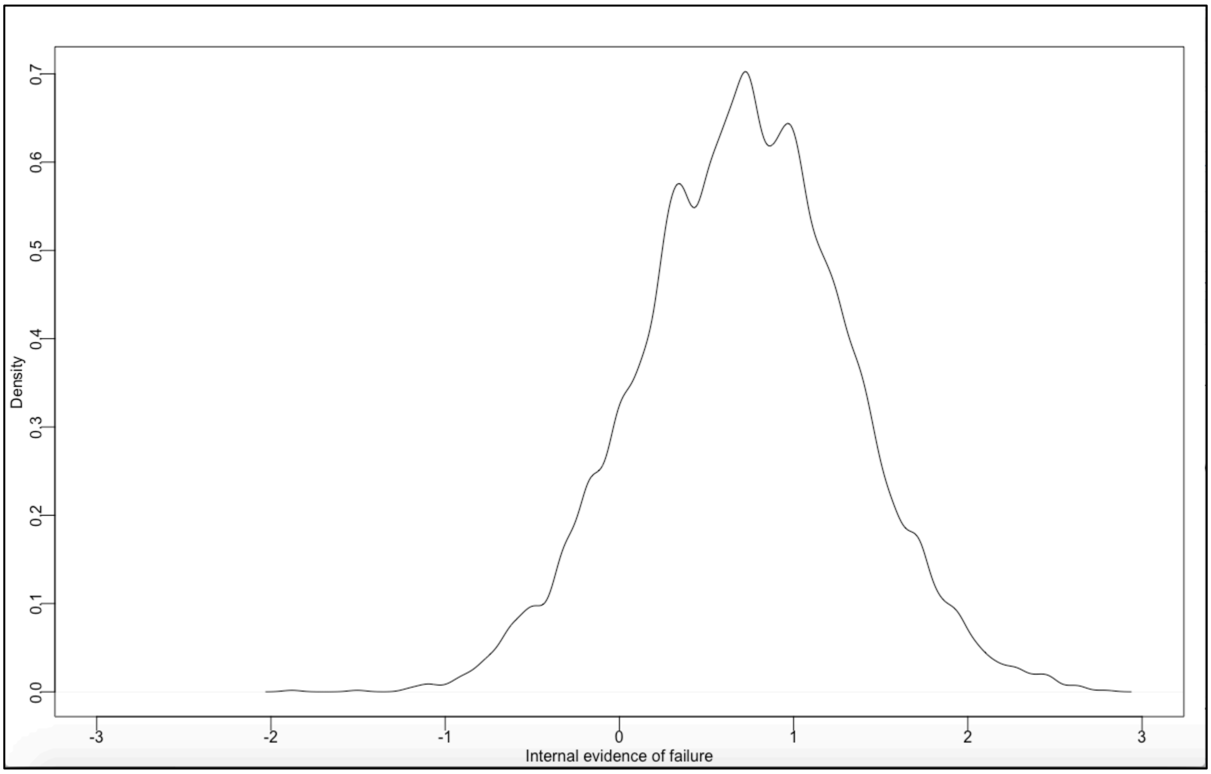

***Figure 42*. Graph showing samples drawn from the posterior distribution produced by Model 14 for the marginal effect of 'internal evidence of failure' on macrostructure similarity scores.**

failure' parameter, as estimated by Model 14, was positive, at 0.71 (SD=0.61) though with a 0.89 HPDI which stretched into negative values: from -0.21 to 1.73.

Figures 43 and 44 illustrate Model 14's predicted effects on macrostructure similarity scores of changing low internal evidence of failure into high internal evidence of failure, when this effect was present in conditions of asocial versus social models, low and high participant age, and low and high attendance to the experimental video. It appears that the change from low to high internal evidence of failure did not cause greater macrostructure similarity scores. In none of the four social model conditions, across Figures 43 and 44's graphs C and D, was there a visibly positive effect of increased internal evidence of failure.

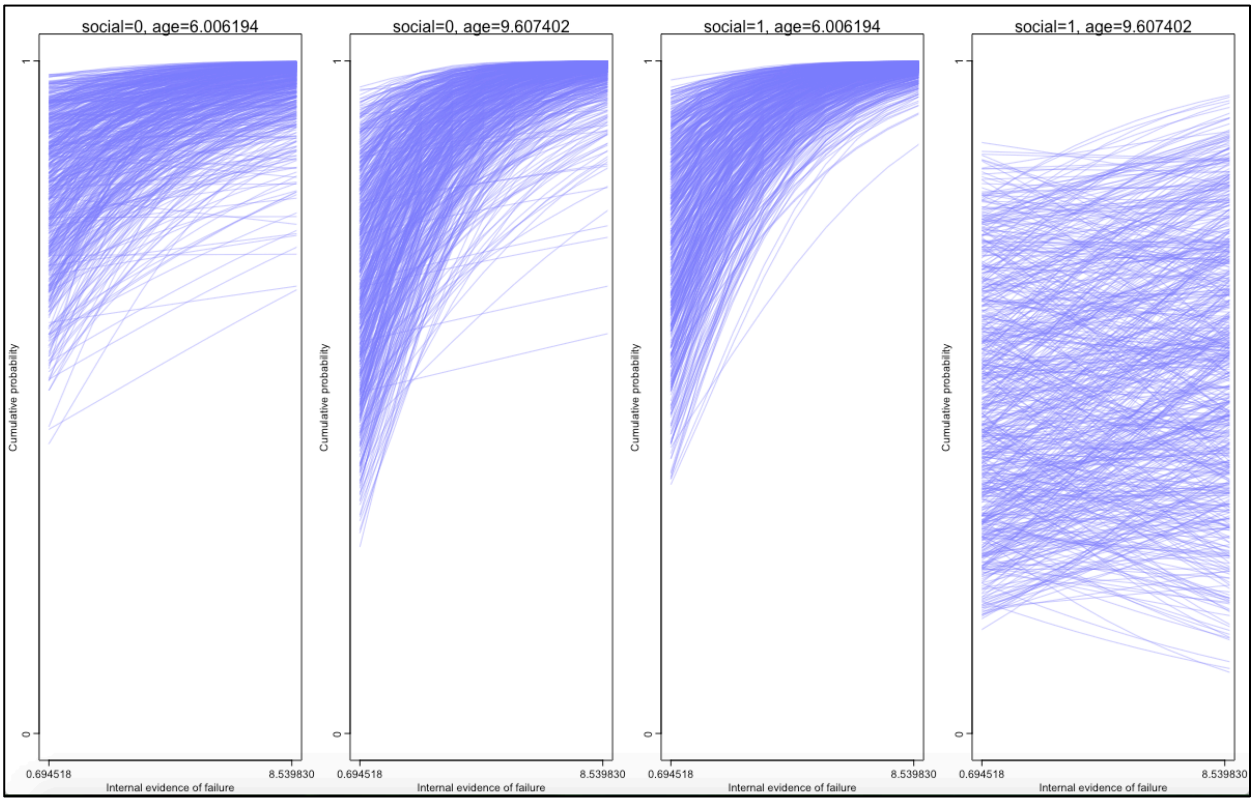

***Figure 43*. Four graphs (from left to right: A, B, C, and D) illustrating Model 14's predicted effects on macrostructure similarity of changing 'low' internal evidence of failure into 'high' internal evidence of failure. 'Low' internal evidence of failure was set at one standard deviation below the mean score for the entire usable dataset, while 'high' was set at one standard deviation above this mean. Graphs A and B (on the left) show the effect of this change when the model was asocial, and graphs C and D (on the right) show the effect of this change when the model was social. Graphs A and C (on the far left and second from right) show the effect of the change in internal evidence of failure when the participant's age was 'low', and graphs B and D (on the second from left and far right) show this effect when the participant's age was 'high'. 'Low' age was set one standard deviation below the mean age of the entire usable dataset, while 'high' age was set one standard deviation above this mean age. For all four of these graphs, the 'attendance to the experimental video' score was set to 2.44: one standard deviation below the mean attendance score for the entire dataset.**

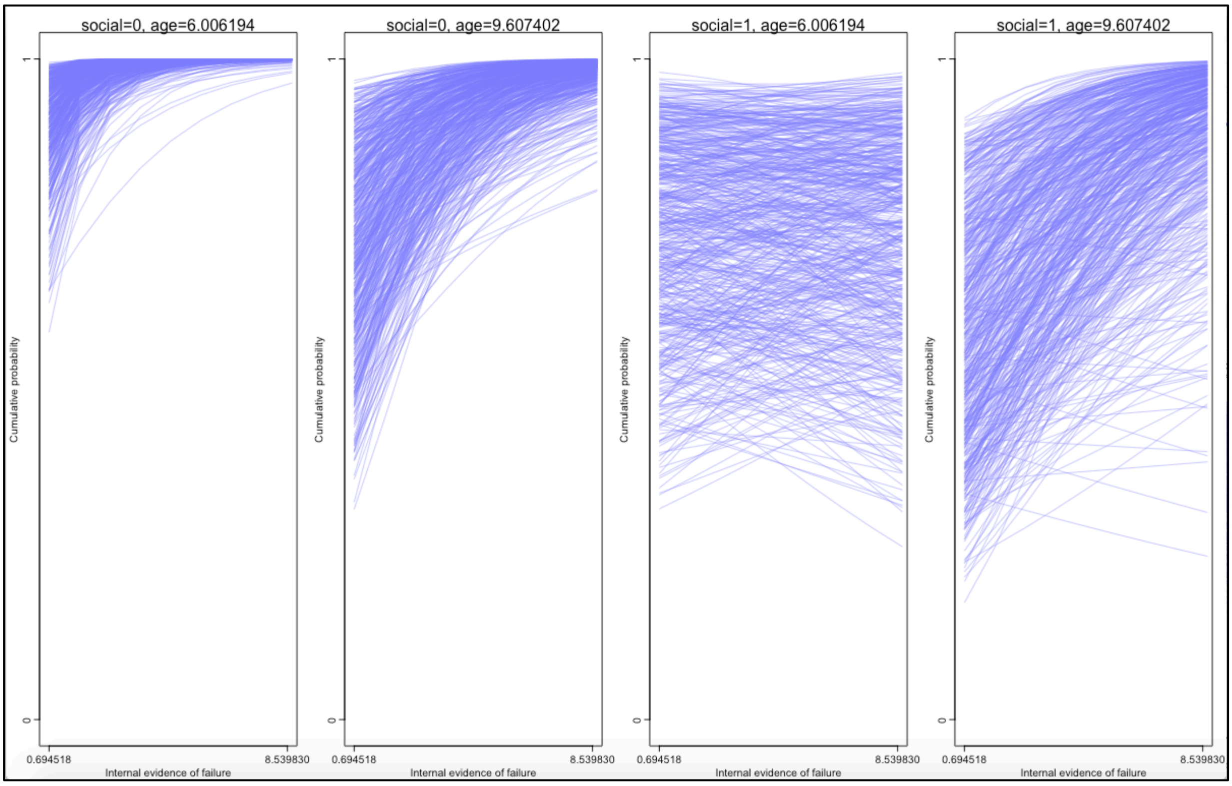

***Figure 44*. Four graphs (from left to right: A, B, C, and D) illustrating Model 14's predicted effects on macrostructure similarity of changing 'low' internal evidence of failure into 'high' internal evidence of failure. 'Low' internal evidence of failure was set at one standard deviation below the mean score for the entire usable dataset, while 'high' was set at one standard deviation above this mean. Graphs A and B (on the left) show the effect of this change when the model was asocial, and graphs C and D (on the right) show the effect of this change when the model was social. Graphs A and C (on the far left and second from right) show the effect of the change in internal evidence of failure when the participant's age was 'low', and graphs B and D (on the far right and second from left) show this effect when the participant's age was 'high'. 'Low' age was set one standard deviation below the mean age of the entire usable dataset, while 'high' age was set one standard deviation above this mean age. For all four of these graphs, the 'attendance to the experimental video' score was set to 11.53: one standard deviation above the mean attendance score for the entire dataset.**

However, this appears to have been more consistently true in the four asocial model conditions, which all demonstrated strong negative relationships between increased internal evidence of failure and macrostructure similarity scores. In fact it appears that a social model in combination with either (a) older children with lower attendance to the video (Figure 43's graph D) or (b) younger children with higher attendance to the video (Figure 44's graph C) reduced the negative effect of changing internal evidence of failure from low to high. Meanwhile, the influence of higher participant age on the effect of higher internal evidence of failure appears to have been complex, as does the influence of participant age. More detailed analyses of the effects of these other variables can be found in Appendix 10.4.

The fourth and final hypothesis explored the effect of internal evidence of failure on variation in macrostructure similarity scores. A key part of defining 'play' is the activity's open-endedness, in which children have greater freedom not only in the means of the activity but in the ends (van Oers 2013). In this playful and unthreatening context, children can explore previously untested solutions to ecological problems (Bateson 2014). When internal evidence of failure is encountered, children may therefore not attempt to replicate the macrostructure of a social model since this would not solve the problem but merely avoid it. If play is about experimentation, children exhibiting higher internal evidence of failure may thus build macrostructure designs which are different from the social model. In Hypothesis 4 I therefore predicted that the rate of children's internal evidence of failure would not be a positive predictor of macrostructure similarity scores. While it is true that none of the conditions showed a positive effect of high internal evidence of failure on macrostructure similarity scores, this cannot be taken as evidence in support of Hypothesis 4 since internal evidence of failure had a visibly stronger negative effect on macrostructure similarity scores when the model was asocial rather than social (the reverse of Hypothesis 3).

The lack of a positive effect of internal evidence of failure when the model was social therefore appears to have been less the product of children 'deciding' not to copy the social model's macrostructure design than it appears higher internal evidence of failure was caused by a macrostructure dissimilar to the social model. This interpretation, however, is not fully in line with the previous evidence from Chapter 5 for the relationship between internal evidence of failure and macrostructure similarity, which indicated that in close-ended tasks, while lower microstructure similarity scores were associated with high internal evidence of failure, higher internal evidence of failure was associated with higher macrostructure similarity scores. From Figures 43 and 44, it appears that with the social model, at least, participant age and attendance to the video interacted to play a key role in the effect of internal evidence of failure on macrostructure similarity scores. In any case, it appears that the internal evidence of failure measure did constitute an interesting source of information for understanding children's copying behaviour in this experiment. Its relationships with microstructure and macrostructure similarity scores indicate that the degree of participant failure in the task was related to the children's building styles and aims. These associations are often best explained by interpreting internal evidence of failure as a product of a child's building style and goal (see Hypothesis 3 also). The question to be answered is, therefore, why children often did not change building styles or goals (for example, copy from the social model) when their building styles and aims seem to have created such increases in internal evidence of failure.

Chapter 7: General discussion, implications for future research, and impact

In this chapter, I synthesise the results from Chapters 4, 5, and 6 to outline the main findings from my thesis. I compare these results with those from previous literature to highlight points of convergence and difference, and use theory introduced in Chapter 2 to suggest explanations for these patterns. This leads into the implications of the thesis for future studies. I conclude the thesis by evaluating (a) its impact on research into children's social learning and (b) the relevance of such a contribution for wider society.

7.1: General discussion

In total, data gave reasonable support to three of the eight hypotheses, leaving five in which the predicted effects were not reliably visible. There is interesting variation in which hypotheses found reasonable support and which did not. A summary of all of the results, from Chapter 4 as well as Chapters 5 and 6 is provided in Table 5. Firstly, Table 5 shows data supported more of Chapter 5's hypotheses than Chapter 6's. While Hypotheses 1, 2, and 3 of Chapter 5 found reasonable support in the data, none of Chapter 6's hypotheses did. Secondly, this pattern may be linked to whether the task was open- versus close-ended. Chapter 5 consisted of three hypotheses about outcome variable variation in the close-ended task, and one hypothesis about outcome variable variation across both open-ended and close-ended tasks. Chapter 6's four hypotheses, meanwhile, all focused on variation in the outcome variable when the task is open-ended. This indicates that the hypotheses were better able to predict the two outcome variables' variation when the task was close-ended rather than open-ended. Thirdly, the pattern may also be linked to whether the outcome variable was microstructure or macrostructure similarity scores. Of four hypotheses predicting variation in microstructure similarity scores, two received reasonable support. Meanwhile, of four hypotheses predicting variation in macrostructure similarity scores, only one was supported by the data. Overall, therefore, there are several results requiring explanation.

Table 5
Summary of the hypotheses and results from Chapters 4, 5, and 6.

Hypothesis	Prediction	Result
Chapter 4: Age	Lower age would be associated with greater copying of the social model	Higher age was associated with greater copying of the social model
Chapter 4: Sex	Females would copy the social modal more than males	Females copied more than males, but only older children showed this consistently
Chapter 4: Attendance to the experimental video	Participants that showed greater attendance to the social video would copy the social model more	There was no consistent positive effect of greater attendance to the video
Chapter 5: Hypothesis 1	The close-ended, rather than open-ended task, would cause participants to build macrostructure designs more similar to the verbal instructions	The close-ended task, but not the open-ended task, was associated with greater macrostructure similarity scores in most conditions
Chapter 5: Hypothesis 2	In the close-ended task, participants would copy the social model when the model was relevant (i.e., social) compared to when the model was irrelevant (i.e., asocial)	Participants built microstructure with greater similarity to the social model when they observed the social model, except amongst younger children who observed the unsuccessful model
Chapter 5: Hypothesis 3	In the close-ended task, participants would copy the successful social model's microstructure more than the unsuccessful social model's microstructure	Older children showed increased copying of the social model's microstructure when the model was more successful. Younger children also showed increases, but no larger than the increase between the unsuccessful and successful irrelevant (i.e., asocial) model
Chapter 5: Hypothesis 4	In the close-ended task, participants would not copy the successful social model's macrostructure more than the unsuccessful social model	When younger children observed the successful social model they showed no increase in the macrostructure similarity of their builds to the social model compared to when the model was asocial. However, older children did show increased macrostructure similarity
Chapter 6: Hypothesis 1	In the open-ended task, participants would copy the successful social model's microstructure more than the unsuccessful social model's	Participants copied the successful social model more than the unsuccessful social model only when the participants themselves exhibited low internal evidence of failure. Even when they did exhibit low internal evidence of failure, the positive effect of social model success was minor
Chapter 6: Hypothesis 2	In the open-ended task, participants would not copy the successful social model's macrostructure more than the unsuccessful social model's	Participants did not consistently copy the unsuccessful social model's macrostructure more than the successful model's. However, neither did they consistently copy the successful model's more than the unsuccessful model's
Chapter 6: Hypothesis 3	In the open-ended task, participants exhibiting greater internal evidence of failure would copy the social model's microstructure more than participants exhibiting less internal evidence of failure	Participants exhibiting greater internal evidence of failure did not copy the social model's microstructure more than participants exhibiting lower internal evidence of failure
Chapter 6: Hypothesis 4	In the open-ended task, participants exhibiting greater internal evidence of failure would not copy the social model's microstructure more than participants exhibiting less internal evidence of failure	Participants exhibiting greater internal evidence of failure did not copy the social model's macrostructure more than participants exhibiting lower internal evidence of failure

I start the discussion with the hypotheses which received reasonable support from the data. Chapter 5's Hypotheses 2 and 3 used data from the close-ended task to test previous findings about the positive effect of (a) a social model and (b) a successful social model on microstructure similarity scores. The results reliably corroborated previous studies such as Smith, Ward and Schumacher (1993), Shalley and Perry-Smith (2001), and Rook (2008) in that children's builds showed increased levels of microstructural similarity to the social model's build when children could observe the social model compared to when children could not observe the social model, when the models were either successful or unsuccessful. These results also reliably corroborated previous findings from Kendal et al. (2005), Carr, Kendal and Flynn (2015), and Turner, Giraldeau and Flynn (2017) that children demonstrate greater microstructure copying when the social model is successful compared to when the social model is unsuccessful. This indicates that the experimental setup and context was able to produce the same behaviours in children that previous social learning experiments could. Other results that contradict expected findings should not therefore be disregarded out of hand. Chapter 5's Hypothesis 1, which predicted that macrostructure designs would display greater similarity to the social model when the task was close- (rarther than open-) ended, also found reasonable support in the data. This indicates that the close- versus open-ended task did have some influence on children's copying behaviour: the close-ended task often caused children's builds to show greater microstructure similarity to the social model regardless of whether the children observed the social model. This provides support for the argument, made in Chapter 2, that close-ended social learning studies constrain variation in children's activities.

I now move on to the hypotheses which did not receive reliable support from the data. Chapter 5's Hypothesis 4 aimed to provide further support to the argument that close-ended conditions constrain children's copying behaviour by predicting that the success of the social model would not be a good predictor of variation in macrostructure similarity scores in the close-ended task. However, model success did have a positive effect on macrostructure similarity scores with older children, who were shown to copy more than younger children in

Chapter 4's analyses. Hypothesis 4 thus appears to have been too tough a test for the constraining effects of the close-ended task; the effect of a successful social model on many children's copying behaviour remained relatively clear when macrostructure was the outcome variable. This indicates that the role of social information remained important for older children's macrostructure building, despite the constraints imposed by the close-ended task (see Hypothesis 1 of Chapter 5).

Chapter 6's four hypotheses were given little support from the data. Given the sample sizes, balance between experimental control and ecological validity, and range of variables data were collected for, these are interesting results deserving of explanation. Hypothesis 1, which predicted that in the open-ended task a successful (rather than unsuccessful) social model would increase children's microstructure similarity scores, gained only the weakest support. In the open-ended task, children only showed greater similarity in their microstructure designs to the successful social model rather than the unsuccessful social model when they also exhibited low internal evidence of failure. Even when this was the case, the positive effect of model success was so weak as to be nearly meaningless; indeed it was no stronger than the positive effect of asocial model success when participants demonstrated high internal evidence of failure. When children did show high internal evidence of failure, the successful (rather than unsuccessful) social model had no discernable effect on microstructure similarity scores.

This is surprising on two counts. (1) High internal evidence of failure is, in the literature, associated with greater reliance on social information (Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a; Caldwell & Millen 2010). This first point is discussed further below, for Chapter 6's Hypotheses 3 and 4. (2) The open-ended task appeared to induce less microstructure copying of the successful social model than in the close-ended task. This means that social model success was associated with weaker similarity to the social model in children's microstructure designs when the task was open-ended rather than close-ended. This could be taken as evidence that the children exhibited less copying behaviour in the open-ended tasks compared with the close-ended tasks, if model success did not induce as great a reliance upon social information when the task was open-ended as it did when the task was close-ended. This maybe fits evidence from creativity research that reliance on social models is exacerbated by stressful conditions (Festinger 1954; cited by Rook 2008), if we assume that the open-ended, playful context induces less anxiety than the close-ended task.

Hypothesis 2, meanwhile, sought to examine whether children in the open-ended task would emulate the macrostructure goal of the unsuccessful social model. This did not find convincing support, as half of conditions with a social model did show some kind of positive effect of model success on macrostructure similarity scores. The results showed considerable variation in children's responses, with examples where the successful model had positive, negative, and neutral effects on macrostructure similarity scores. There was therefore no evidence that a successful social model either reliably increased or reliably decreased children's copying of macrostructure when the task was open-ended. This was a different result from Chapter 5's Hypothesis 4, which examined children's macrostructure copying in response to a successful model but close-ended task. This result therefore supports the interpretation above of Chapter 6's Hypothesis 1, and indicates that in the open-ended task children were not incentivised to copy the successful social model more than the unsuccessful social model as they were in the close-ended task. This pattern also appears related to Sheridan et al.'s (2016) findings that children, given a more open-ended context, exhibit higher rates of innovation than reported in previous studies.

Chapter 6's third and fourth hypotheses show essentially no support from the data. While Hypothesis 3 posited that participants exhibiting higher internal evidence of failure should rely on social information about microstructure design to a greater extent than participants encountering less of a challenge in the task, the results indicated the opposite: that participants exhibiting higher internal evidence of failure copied the model's microstructure design less.

Hypothesis 4 predicted that high internal evidence of failure would be associated with high macrostructure similarity scores when the model was social. However, when the model was social, I found a weakly negative relationship between high internal evidence of failure and macrostructure similarity, and this was less negative than the same relationship when the model was asocial. The likely cause of these results is that participants who built a structure less like the model's structure encountered greater instances of collapse. However, this does not explain why, when they encountered this internal evidence of failure, they did not address this evidence of failure by copying the model. An explanation could be that in the open-ended task there was no incentive to avoid internal evidence of failure, since there was no external goal which participants failed to achieve. The explanation, in the literature, for reliance on social information when internal evidence of failure is encountered is that the relative risks of the social information not being useful are reduced (Feldman, Aoki & Kumm 1996; Boyd & Richerson 1985). The studies in which this effect is documented employ close-ended tasks in which it is obvious when a participants does not succeed (e.g., Caldwell & Millen 2010; Williamson, Meltzoff & Markman 2008; Wood, Kendal & Flynn 2013a). In the open-ended context employed in the present study, such reliance on social information was not incentivised in this way. This open-ended context, in which success was not incentivised, is in line with definitions of play in Chapter 2, which describe it as a safe space for experimentation in which there are no ramifications of failure (Bateson & Martin 2013; Gopnik 2012; Cook, Goodman & Schultz 2011). The data therefore perhaps suggest that copying is less widespread in open-ended and playful conditions, since children's internal evidence of failure does not motivate copying behaviour.

In Rook's (2008) open-ended task, however, adult participants were observed to increase their reliance on copying when they found the task harder and when the model performed better. This perhaps suggests that the effects of open-ended, playful conditions reported here might be limited to children. An argument could also be made that these results were caused by a task in which causation was transparent, and by participants' likely prior familiarity with building towers from blocks (Want & Harris 2002). Neldner, Mushin and Nielsen (2017), for instance, find that the degree of task opacity can inhibit the ability of children to produce innovative solutions in an asocial task, and Whalley, Cutting and Beck (2017) find evidence of improved performance in an asocial task when children had prior experience with the task. These arguments do not diminish the significance of the results reported, however, since greater signs of children's copying of the social model were evident when the task was close-ended, meaning that the difference between the close- and open-ended conditions remains to be explained. Nevertheless, it cannot be ruled out that these specific properties of the building block task may have been factors in permitting such a difference in children's copying between the close- and open-ended conditions.

The data also exhibited interesting results about the role of participant age on copying behaviour. The effect of age is particularly interesting as it was the reverse of some previous findings. Carr (2016) found that younger children were more faithful copiers than older children, while Walker and Andrade (1996) found that younger children displayed greater social conformity compared with older children. The present data, meanwhile, show that it was older children who copied the social model to the greater extent. Furthermore, this effect held true across both open- and close-ended conditions, and across microstructure and macrostructure outcome variables. In the present data, therefore, it was the younger children who demonstrated build structures which were more different from the social model, and this difference between younger and older children's similarity scores was much clearer when they observed the social rather than asocial model. One explanation for this effect may be 'functional fixedness' (German & Defeyter 2000; Carr 2016), in which knowledge of an object's function reduces a child's ability to repurpose it for a different use, the effect of which is known to increase from the age of five years old (Defeyter, Avons & German 2007). This could lead to younger children adapting the social model's information to a greater degree than older children, causing their builds to resemble the social model's less. Older children have also been found to demonstrate increasing 'over-imitation' when presented with causally opaque tasks (DiYanny, Nini & Rheel 2011; Flynn & Whiten 2008), or when tasks are framed as being normative rather than instrumental (or communicative; Clay, Over & Tennie 2018). However, this would not appear to explain the present findings from a causally transparent building blocks task with both open- and close-ended instruction. Furthermore, the setup of the experiment was designed to provide little which could be interpreted as a normative cue.

The data also indicated a sex difference in older but not younger children's builds. Across both microstructure and macrostructure similarity outcome variables, all but one condition showed that older females generally built structures more similar to the model (when they observed the social model) than male participants. When the model was asocial, females did not show reliably higher microstructure and macrostructure similarity scores. This indicates that the effect of being female is in copying behaviour rather than simply building style. This effect is, however, much messier among the younger children. This effect of being female is broader than that found by Brand, Brown and Cross (2018), who observed it only in participants' use of social information when asocial learning was riskier. The association of females with greater copying behaviour appears to tally with some evidence from apes that males exhibit greater innovative, 'asocial' learning than females (Reader & Laland 2001; Vale et al. 2017b; and Ervin et al. 2015's rodents). However, an explanation drawn from differences in what males and females are expected to play with and how they play (see Freeman 2007; Brahms & Crowley 2016; see also Ehrlinger & Dunning 2003) should not be overlooked. Nevertheless, while participant age carried enough predictive value to be included in several of the models used for data analysis (e.g., Model 7 on page 74, and Model 14 on page 113), participant sex did not carry enough predictive value to be included in any models apart from the two directly addressing its effects on microstructure and macrostructure similarity scores (Models 3 and 4 in Chapter 4). This suggests that participant sex does not play a significant role in understanding variation in either microstructure or macrostructure similarity scores according to any of the other factors tested.

A predominant theme throughout the results is variation and interdependence. Even where the hypothesised effects of a predictor variable were supported by data, these effects varied in strength, as well as sometimes direction, dependent on the status of other variables included in each statistical model. Even strong positive effects of variables, such as social model success on macrostructure similarity scores (Figure 39, graphs C and D; page 105), could be reversed by changing, for example, internal evidence of failure from one standard deviation below the mean to one standard deviation above the mean. The observable variation in the results indicate that children's copying behaviour was contingent on several factors, including the open- versus close-ended task, the social versus asocial model, the success of both the social and asocial model, degree of internal evidence of failure, as well as participant age, sex, and attendance to the experimental video (as included in Model 12 on page 102). Furthermore, there are slight differences between how children copied microstructure versus macrostructure under the same experimental conditions. For example, model success was a slightly more consistent positive predictor of variation in microstructure similarity scores than macrostructure similarity scores in the open-ended task, as was the case also in the close-ended task (see Flynn & Whiten 2008 for a related finding in a complex puzzle box task). All of this contributes to arguments made in the social learning literature that children's copying is context-sensitive and flexible rather than indiscriminate (Over & Carpenter 2012; Evans et al. 2017; Legare et al. 2015; Kendal et al. 2018; see also Whitehead & Richerson 2009).

7.2: Implications for future research

A key implication of the present study for future research is that more attention must be paid to the ecological validity of the experimental task. This is because the data presented here show that children's copying behaviour can differ between tasks which are more close-ended and tasks which are more open-ended. This builds on Sheridan et al.'s (2016) finding that the rate of children's innovation is increased by an open-ended, creative context. Therefore, experimental setups which focus only on children's copying in a close-ended task should be cautious in applying their findings to children's copying outside of such close-ended contexts, and especially to children's copying in open-ended or playful contexts. Indeed it appears that amongst contemporary hunter-gatherer groups, adults (and older children) limit their interventions into children's learning in favour of greater reliance on autonomous and child-led exploration (Boyette & Hewlett 2018; also Lancy 2017). More research would therefore be useful in comparing children's copying behaviours across various combinations of experimental conditions, whilst systematically varying the extent to which the task is open- versus close-ended. This would be important in understanding the extent to which the differences between children's copying in close-ended and open-ended conditions were made possible by the causally transparent and already familiar building blocks task.

In fact, the present dataset could be used to explore this question further. There are comparisons between open-ended and close-ended conditions which were not analysed here. For example, data already exist to analyse the influence of internal evidence of failure on children's copying with a close-ended task. This analysis would be important in testing whether the cause of children's lack of copying in conditions of high internal evidence of failure is due to the open-endedness of the task. Furthermore, data already exist to perform the analysis of how children's copying, in the open-ended task, is affected by the 'social' versus 'asocial' model. The results of Chapter 5's Hypothesis 2 show that, in the close-ended task, children exposed to the social model displayed greater microstructure similarity scores. Comparisons between Chapter 5's Hypotheses 3 and 4 with Chapter 6's Hypotheses 1 and 2 already indicate that children's microstructure and macrostructure copying reliably increases when the social model is successful (rather than unsuccessful) when the task is close-ended but not when it is open-ended. To investigate whether this is true also of a comparison between children's exposure to the social versus asocial model would be key in understanding how deeply children's copying is affected by the close- versus open-endedness of the task.

A limitation of the present research is that I did not consider individual-level psychological differences, which are becoming increasingly recognised in the social learning literature (Rawlings, Flynn & Kendal 2017). A strength of Rook's (2008) study is that some psychological parameters are included in their analyses. Their work suggests individual-level variation is present in participants' reliance on model exemplars. Rook (2008) specifically focuses on differences in individuals' self-regulatory focus towards either risk prevention or reward promotion, and differences in individuals' attention to either specific or general information. The former of these parameters appears likely related to how children respond to internal evidence of failure in their own building, and the latter seems relevant for the distinction between microstructure and macrostructure copying. Nevertheless, while such differences in psychological disposition may be interesting avenues for further research, I do not believe that they threaten the reliability of the present study's results, since, as Rook and van Knippenberg (2011) write, model exemplars do incite copying from both promotion-focused and prevention-focused individuals. Such individual-level psychological variation thus may be more a refinement of than a problem for the present thesis.

Another avenue for further research will be to analyse with greater precision the activities of the child in the context of play building. This approach would ask questions about the ways in which children approach choices of what and when to copy a model. Work on action phases and mindset theory (Moreau & Engeset 2016; Gollwitzer 2012; Lewin et al. 1944) indicates that there may be different cognitive processes involved in how people respond to problems of goal-setting and goal-striving. These may be useful in better understanding the difference between copying macrostructure versus microstructure. Furthermore, the present study assumes that children do work towards a preconceived macrostructural goal, which is not an uncontroversial assumption. Research in embodied cognitive science, or in situated cognition (e.g., Brown, Collins & Daguid 1989), criticises this representationalist account of behaviour (see Wilson & Golonka 2013; Dove 2015; Barrett, Henzi & Lusseau 2012; influenced by the phenomenological tradition of Merleau-Ponty 1962). Wilson and Golonka (2013) argue it is problematic to assume disembodied 'states of the mind' which account for the behaviour of an agent, and that instead any given behaviour can be considered the result of interactions between a cognitive system and the environment it is embodied within. Future work, which does more acutely analyse children's copying behaviour, might therefore attempt to reconfigure the microstructure-macrostructure distinction to recognise both levels as emergent properties of children's interaction with the building stimulus, and to study copying behaviour as a solution to a task (defined from a first-person perspective) in which the agent assembles a given set of resources (spanning brain, body, and environment) into a system capable of solving the task (also see citations 28 to 33 in Sheridan et al. 2016). Such a situated account of the cognitive processes of copying in building would fit well with the activity theory approach of van Oers (2013) already applied in the present thesis to describe the developmental niche of play. See Flynn et al. (2013) for a synthesis of situated cognition with activity theory, niche construction, and developmental psychology.

7.3: Impact of current research

I now briefly summarise the impact of the current research on the social learning literature. I created experimental conditions which succeeded in replicating prior experiments (Smith, Ward and Schumacher 1993; Shalley and Perry-Smith 2001; Rook 2008; Kendal et al. 2005; Carr, Kendal and Flynn 2015; Turner, Giraldeau and Flynn 2017) which found that children, in a close-ended task, are incentivised to copy microstructure design when a social model is present and when the social model is successful rather than unsuccessful. I found that the macrostructural design of children's builds was constrained by the close-ended task (compared with a more open-ended task in which no goal or end-state was verbally instructed). However, I also found that there was still a positive effect on copying the social model's macrostructure when that model was successful rather than unsuccessful.

Contrary to expectations, data indicated that in the open-ended task, a successful social model only increased children's copying of microstructure negligibly when the child encountered low internal evidence of failure, and data did not demonstrate evidence that, in the open-ended task, a successful social model reliably increased children's copying of macrostructure. Further contrary to expectations, data from the open-ended task showed no positive effects of higher internal evidence of failure on children's copying of either microstructure or macrostructure. The suggested explanation for these results is that, for children at least (see Rook 2008 for adults), copying is not incentivised in playful conditions as it is in close-ended goal-focused conditions. This supports research into play as social niche for children's development (Sheridan et al. 2016; Riede et al. 2018; citing Palagi, Stanyon & Demuru 2015) which suggests that open-ended play provides a context in which children do not display such conservatism towards social information as found in close-ended experimental setups.

This suggestion requires further experimental work to understand the present results. Firstly, research must test that it is specifically children interacting with an open-ended, playful task, rather than a close-ended task, that is the cause of the findings. Secondly, research must test the role of other factors, such as the lack of causal opacity in the task and children's prior experience with building blocks, which may be important in permitting the difference observed in children's copying between close- and open-ended conditions. The interaction of 'over-imitation' with the difference between close-ended and open-ended tasks thus remains to be explored. Furthermore, this difference in copying between open-ended versus close-ended tasks may be especially relevant for studying children's copying in 'WEIRD' contexts, since an absence of formal schooling systems (see Sterelny 2012) may make open-ended tasks more abundant in, and perhaps more significant to, children's development (Lancy 2017; Boyette & Hewlett 2018).

It is widely considered today that the impact of research should not be restrained to the academic field, but instead should be felt in the lives of people outside the discipline (Rudman et al. 2017; citing HEFC, SFC, HEFCW, & DELNI 2012). Research on topics related to child development, learning, and play have often found usefulness in informing educational policy and practice (Siegler 2016; also see Frost, Wortham & Reifel 2012; Golinkoff, Hirsh-Pasek & Singer

2006). This is also possible for the current findings: interest may be taken in the apparently lesser importance of copying for children exposed to either external or internal evidence of failure in the open-ended task. This could have implications for under what contexts it is useful to introduce model exemplars for children to copy from, and when to leave children to experiment by themselves without exemplars. It is vital to recognise, however, that in discussing impact on wider society the current data do not have the last say, and that rash confidence in these findings without careful consideration of their place within the history of the discipline would not benefit wider society (see McElreath 2016).

One way in which the current study makes tangible impact is in bringing the scientific process out of the laboratory (see Rudman et al. 2017). I argued in Chapter 3 that this balance between experimental control and 'real-world' setting allows recruitment of participants who would not otherwise have taken part in research. Several guardians indeed said that their children were excited to have taken part in 'real-life' research. More quantitatively, I asked each participant's guardian whether they would like to receive information about the study's final results. Of 580 guardians asked, 425 (73.28%) gave an email address to receive more information about children's social learning in play. A lay summary of the thesis, which the 425 guardians will have access to, can be found in Appendix 11. This helps to validate the research topic as one in which non-specialists have interest, and to validate the research as a mutually beneficial exercise in which scientific enquiry can progress at the same time as enabling research transparency and greater scientific understanding amongst a non-specialist audience. I therefore conclude that this thesis has impact for both specialist and non-specialist audiences, in furthering scientific understandings of children's social learning in the context of play and in expanding these scientific understandings beyond the laboratory.

Appendix 1: Information and consent form examples (Centre for Life)

Department of Anthropology
Durham University
DH1 3LE
Contact Name: Guy Lavender Forsyth
Mobile: 07985 144078
Email: g.a.lavender-forsyth@durham.ac.uk
Supervisors: Dr Rachel Kendal, Dr Jeremy Kendal

Dear Parent/Guardian,

This is an information and consent form about your child's participation in a study that I (Guy Lavender Forsyth) am running at the Centre for Life.

My project's title is '*Flexibility in social and asocial learning*' (supervised by Drs Rachel and Jeremy Kendal). I am investigating how children use different information when they learn under different conditions. I want to find out if they learn differently when they can play freely compared to when they are given a task to achieve, and whether their experiences in play and in observing a model to learn from change what sorts of information they copy.

Your child's participation will take approximately 5 minutes. It involves them playing at a permanent exhibit here in the Brain Zone. The exhibit has been designed to be enjoyable for the children. The children can decide to leave the study for any reason, at any time.

I will videotape the children as they build so that later on I can record the ways in which they used the building blocks. These videos will be used only for this purpose. Data collected from the children will be anonymised. All recordings made will be stored in secure locations, and will be destroyed at the end of my study.

I have undergone a full DBS (Disclosure and Barring Service) check, and have been approved for working with children. I have worked with children in the past as a researcher in the Centre for Life, and Durham researchers have a positive record of research in collaboration with the Centre for Life.

If you are willing for your child to participate in my study please complete the consent form below. If you have any further questions about my study I would be glad to answer them via email.

Many thanks,

Guy Lavender Forsyth (postgraduate, Durham University)

Consent form: ***Flexibility in social and asocial learning***

Parent/Guardian, please complete these questions below

1) Child's name (to identify which children have been given consent):

2) Child's date of birth (day/month/year):

3) Child's sex (please circle):

Female Male Other Prefer not to say

4) Please choose whether you consent to your child's participation in my study (please circle)

I consent I do not consent

5) Signature: Date:

6) Do you want to be sent a summary of my study's results (please circle)?

Yes No

(If 'yes', please provide an email address)

Email:

Appendix 2: Photograph of the experimental setup

Appendix 3: Microstructure coding procedure

Microstructure coding

Below is a list of characters which children's builds may exhibit. When coding the presence of these characters in a build, this serves as a checklist to see which characters a child has used, by noting each character's presence or absence. Coding therefore does not include frequency counts of any of the characters.

1. Block lying on thin long side
2. Block lying on thin short side
3. Block lying on broad side

4. Block incline on edge between thin long and broad sides
5. Block incline on edge between thin short and broad sides
6. Block incline on edge between thin long and thin short sides

7. 90° overlap of a block with a block directly above
8. 180° overlap of a block with a block directly above
9. Diagonal overlap of a block with a block directly above

10. 90° articulation of a block to a block on the same plane
11. 180° articulation of a block to a block on the same plane
12. Diagonal articulation of a block to a block on the same plane

13. A block with between 1 and 3 blocks resting directly above it
14. A block with between 3 and 7 blocks directly above it
15. A block with between 7 and 15 blocks directly above (15 being the maximum number of blocks – on their thin edges – which can lie on a block on its broad edge)

16. 2 blocks, neither more nor less, touching the table

Notes:

- This scheme comprises five groups of three possibilities each, and one extra group of binary presence/absence.
- The first five groups each have three possibilities because these possibilities are not mutually exclusive. So one build can plausibly have all three possibilities of block orientation (e.g., lying on thin long edge, lying on thin short edge, and lying on broad edge). Therefore, to stop any one group outnumbering the rest, each of these groups has a maximum of three possibilities.
- The one extra group, however, contains options which are mutually exclusive (i.e., if a build has one option, it does not have any of the other options of that group). A participant build either has the same

number of blocks touching the table of the model build or a different number.

- All of the possibilities are positive rather than negative. Each possibility is an attribute that a build can have, rather than the absence of an attribute.
- The 'diagonal' articulation and overlap of blocks is not a category for blocks which deviate slightly from perfect 180' or 90'. A diagonally oriented block will stand out as such.
- The 'block incline' category includes blocks which are on a slope in the direction of the edge pointing downwards but where this edge is not actually touching the table/another block.
- If a block is inclined along the 'thin long–broad' axis, depending on the degree of incline, the block could either be in the 'broad' or 'thin long' position. If the block is angled more vertically than diagonally, this would be the 'thin long' position, if more horizontally than diagonally, this would be the 'broad' position. In most cases anyway, a build has either one or the other type, rather than both, with both types equally different from the model build.
- Most of the time, the limits of a build's structure are clear. However, the precise boundary of a child's build can be debatable. In this case, use common sense and intuition to define the boundary of building.
- Occasionally, when images are taken from the video recordings, they do not picture the build's base. When this is the case, I checked the corresponding video to see how the base was built.
- 'On the same plane' means when two or more blocks are lying on a surface of the same height. This is regardless of the position of the block – the plane is the surface underneath it.

Example of microstructure similarity coding:

Build	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Differences (frequency)**	**Similarity score**
Model	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	N/A	N/A
Participant 1 (SC)	1	0	1	0	0	0	1	1	0	0	1	0	1	0	0	0	3	13
Participant 2 (UO)	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	3	13
Participant 3 (ASO)	1	1	1	0	0	0	0	1	0	1	1	1	1	0	0	0	7	9

The numbers in bold at the top of the columns correspond to the list of 16 microstructure characters.
Differences between the example participant and model builds are highlighted in yellow.

Appendix 4: Information and consent form examples (Prolific Academic)

I am Guy Lavender Forsyth, a Masters student from the Department of Anthropology at Durham University. I am researching building behaviour in children.

I have asked individual children to build structures from blocks. I would like to know how they build, specfically whether copying influences their building behaviour.

To do that I need your help. You will be asked to rate pairs of block structures based on how similar the structures in the pictures are to each other. You need to focus on similarity between the overall shapes of the structures, since details of how specific blocks are put together are not important. You will have an option to answer on a scale from 1 (very different) to 7 (very similar).

The rating will take approximately 15 minutes. Please answer based on your first impression and do not think to much about how similar the structures are. Please complete this study in one go.

The data consist of your responses (numbers from 1 to 7), consent, age, gender, the date and time. If you have any more questions about my study you can reach me through email: g.a.lavender-forsyth@durham.ac.uk.

Press any key to begin.

Consent form

I consent to participate in this session, which will involve 130 pairs of pictures I will have to rate by similarity on the Likert scale from 1 (very similar) to 7 (very different). Â

I understand that all data will be kept confidential by the researcher. My personal information will not be stored with the data.Â I am free to withdraw at any time without giving a reason. Â

I consent to the publication of study results as long as the information is anonymous so that no identification of participants can be made.Â Â

The study has received approval from the Research Ethics Committee by Department of Anthropology of the University ofÂ Durham. Â Â

☐ I have read and understand the explanations and I voluntarily consent to participate in this study.

Start

Appendix 5: Model descriptions and comparisons for Chapter 4

5.1: Predicting microstructure similarity scores by participant age

Model 0.1, below, used 'age' (G) as the sole predictor of variation in microstructure similarity scores:

(Model 0.1)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_G G_i$

The model below added the variable 'female' to Model 0.1:

(Model 0.2)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_F F_i + \beta_{GF} GF_i$

A comparison of the two models revealed that Model 0.2 took only 0.19 of the Akaike weight compared to the 0.89 of Model 0.1. Model 0.1 had a WAIC value 2.9 units lower than that of Model 0.2, the standard deviation of this difference being 2.55. That the standard deviation of the difference was smaller than the size of the difference itself was relatively strong evidence that one model made better predictions than another. Model 0.3, below, thus added another variable ('attendance to the video', T) to Model 0.1, rather than Model 0.2.

(Model 0.3)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_T T_i + \beta_{GT} GT_i$

Compared to Model 0.1, Model 0.3 gained only 0.13 of the weight. Its WAIC score was 3.8 units lower than that of Model 0.1 (SD=1.54). Thus Model 0.4, below, added the variable 'open' to Model 0.1.

(Model 0.4)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_{GO} GO_i$

Model 0.4 improved on Model 0.1, by taking 100% of Akaike weight. The WAIC difference between the two models was 22.5 (SD=10.97). Model 0.5, below, added interactions with the variable 'social' (S) to Model 0.4.

(Model 0.5)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GOS} GOS_i$

This model was another improvement. Compared to Model 0.4, Model 0.5 took 100% of the Akaike weight. The difference in WAIC values was 54.3 (SD=16.14). Model 0.6, below, thus added the variable 'successful' to Model 0.5.

(Model 0.6)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GU} GU_i + \beta_{GOS} GOS_i + \beta_{GOU} GOU_i + \beta_{GSU} GSU_i + \beta_{GOSU} GOSU_i$$

Model 0.6 was not an improvement on Model 0.5. Model 0.6 took only 0.08 of the weight. The difference in WAIC scores was 4.9, with a standard deviation of 4.38. I therefore added the last variable to be considered, internal evidence of failure (N), to Model 0.5.

(Model 0.7)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GN} GN_i + \beta_{GOS} GOS_i + \beta_{GON} GON_i + \beta_{GSN} GSN_i + \beta_{GOSN} GOSN_i$$

This model did improve on Model 0.5. In a comparison between the two, Model 0.7 took 0.98 of the weight, the difference in WAIC scores between the two being 7.9, albeit with a standard deviation of 10.18. Model 0.7 is therefore the model used for assessing the influence of participant age on microstructure similarity scores.

5.2: Predicting macrostructure similarity scores by participant age

Model 0.8, below, used 'age' (G) as the sole predictor of variation in macrostructure similarity scores:

(Model 0.8)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i$$

The model below added the variable 'female' to Model 0.8:

(Model 0.9)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_F F_i + \beta_{GF} GF_i$$

Model 0.9 made no improvement on Model 0.8, with an Akaike weight of 0.16 to Model 0.9's 0.84. The difference between the WAIC scores was 3.2 (SD=1.89).

Model 0.10, below, thus added the 'attendance to the video' variable to Model 0.8.

(Model 0.10)

A_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_G G_i + \beta_T T_i + \beta_{GT} GT_i$

Again Model 0.8 was not improved upon. Model 0.10 took 0.15 of the weight, with a WAIC score 3.5 units higher than that of Model 3.8 (SD=1.19). Model 0.11 added 'open' to Model 0.8.

(Model 0.11)

A_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_G G_i + \beta_O O_i + \beta_{GO} GO_i$

The comparison measures indicated greater likelihood that Model 0.11 provides more useful predictions about the data than Model 0.8. Model 0.11 achieves the higher Akaike weight, of 0.72, and the lower WAIC value, by 1.9 units, though the standard deviation of this difference was 5.23. To side with the higher probability, I continued with Model 0.11, to which the 'social' variable was added below.

(Model 0.12)

A_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GOS} GOS_i$

This model was again an improvement. Model 0.12 took 100% of the weight in relation to Model 0.11, with a WAIC difference of 37 (SD=13.4). Model 0.13, below, added 'successful' (U) to Model 0.12.

(Model 0.13)

A_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GU} GU_i + \beta_{GOS} GOS_i +$
$\beta_{GOU} GOU_i + \beta_{GSU} GSU_i + \beta_{GOSU} GOSU_i$

The addition of a variable for model success did not improve model predictions. In comparison with Model 0.12, Model 0.13 took just 0.03 of the weight, and had a higher WAIC score by 7 units (SD=5.07). Model 0.14, below, added 'internal evidence of failure' (N) to Model 0.12.

(Model 0.14)

A_i~Ordered($\mathbf{p}$)

$$\text{logit}(p_k)=\alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GN} GN_i + \beta_{GOS} GOS_i + \beta_{GON} GON_i + \beta_{GSN} GSN_i + \beta_{GOSN} GOSN_i$$

This was again an improvement. Compared with Model 0.12, Model 0.14 took all of the Akaike weight. The WAIC difference between the two was 24.3 (SD=13.02).

5.3: Predicting microstructure similarity scores by participant age

Model 0.15, below, used the variable 'female' to predict variation in microstructure similarity scores.

(Model 0.15)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i$$

The model below added the variable 'age' to Model 0.15:

(Model 0.16)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_{FG} FG_i$$

This model did improve on Model 0.15. Model 0.16 took 100% of the Akaike weight, the difference in WAIC scores being 14.8 (SD=9.09). Model 0.17, below, thus added a variable for participant attendance to the video (T) to Model 0.16.

(Model 0.17)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_T T_i + \beta_{FG} FG_i + \beta_{FT} FT_i + \beta_{FGT} FGT_i$$

Model 0.17 made no improvement on Model 0.16, which gained 0.96 of the weight in a comparison between the two. The difference between the WAIC scores of the two was 6.5 (SD=1.35). Therefore 'open' (O) was added to Model 0.16 rather than Model 0.17.

(Model 0.18)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FGO} FGO_i$$

This model was an improvement on Model 0.16, compared with which it achieves 100% of the weight. The WAIC difference between the two was 22.2 (SD=10.93). I thus added 'social' (S) to Model 0.18.

(Model 0.19)

I_i~Ordered($\mathbf{p}$)

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FOS} FOS_i + \beta_{FGOS} FGOS_i$$

This model was again an improvement. Compared with Model 0.18, Model 0.19 took 100% of the weight, the difference in WAIC values being 46.8 (SD=16.17). The next variable to be added was the one for model success (U).

(Model 0.20)

I_i~Ordered($\mathbf{p}$)

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FU} FU_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGU} FGU_i + \beta_{FOS} FOS_i + \beta_{FOU} FOU_i + \beta_{FSU} FSU_i + \beta_{FGOS} FGOS_i + \beta_{FGOU} FGOU_i + \beta_{FGSU} FGSU_i + \beta_{FOSU} FOSU_i + \beta_{FGOSU} FGOSU_i$$

Compared with Model 0.19, Model 0.20 gained none of the Akaike weight. The difference in the WAIC values of the two models was 14.5 (SD=4.05). Model 0.21, below, thus swapped the variable for model success with one for participant internal evidence of failure (N).

(Model 0.21)

I_i~Ordered($\mathbf{p}$)

$$\text{logit}(p_k)=\alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FN} FN_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGN} FGN_i + \beta_{FOS} FOS_i + \beta_{FON} FON_i + \beta_{FSN} FSN_i + \beta_{FGOS} FGOS_i + \beta_{FGON} FGON_i + \beta_{FGSN} FGSN_i + \beta_{FOSN} FOSN_i + \beta_{FGOSN} FGOSN_i$$

This model did take the greater part of the Akaike weight in a comparison with Model 0.19. However the difference was narrow: with Model 0.19 on 0.43 and Model 0.21 on 0.57. The difference between their WAIC values was just 0.6, with a standard deviation of 9.46. Essentially there was no difference between the models, since the difference calculated between the two was overwhelmed by the degree of uncertainty in the calculation. However, it was notable that Model 0.21 achieved as good a WAIC value as Model 0.19 despite the risk of overfitting with so many parameters. It was therefore with this model that I graphed the effects of participant sex on microstructure similarity scores.

5.4: Predicting macrostructure similarity scores by participant sex

Model 0.22, below, used the variable ‘female’ (F) to predict variation in macrostructure similarity scores.

(Model 0.22)

$A_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_F F_i$

The model below added the variable ‘age’ to Model 0.22:

(Model 0.23)

$A_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_{FG} FG_i$

Model 0.23 took all of the weight, with a WAIC difference of 38.5 (SD=12.65). I thus added ‘attendance to the video’ (T) to Model 0.23.

(Model 0.24)

$A_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_T T_i + \beta_{FG} FG_i + \beta_{FT} FT_i + \beta_{FGT} FGT_i$

This model was not an improvement on Model 0.23, in comparison with which it gained only 0.07 of the Akaike weight. The WAIC score difference between the two models was 5.2 (SD=2.69). The next variable, ‘open’ (O), was added to Model 0.23.

(Model 0.25)

$A_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FGO} FGO_i$

There was greater probability that Model 0.25 improved on Model 0.23 than that was did not, gaining 0.73 of the Akaike weight. The difference between the two models’ WAIC scores was 2, with a standard deviation of 6.34. To side with the greater probability, I continued with Model 0.25. The next model added interactions with ‘social’ (S).

(Model 0.26)

$A_i \sim \text{Ordered}(\mathbf{p})$

$$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FOS} FOS_i + \beta_{FGOS} FGOS_i$$

This addition did resulted in an improved model, with Model 0.26 taking 100% of the weight, and a WAIC score 35.6 units lower than that of Model 0.25 (*SD*=13.31). I next added the variable 'successful'.

(Model 0.27)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = \alpha_k &+ \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FU} FU_i \\ &+ \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGU} FGU_i + \beta_{FOS} FOS_i + \beta_{FOU} FOU_i + \beta_{FSU} FSU_i + \\ &\beta_{FGOS} FGOS_i + \beta_{FGOU} FGOU_i + \beta_{FGSU} FGSU_i + \beta_{FOSU} FOSU_i + \beta_{FGOSU} FGOSU_i\end{aligned}$$

This model gained no Akaike weight relative to Model 0.27. The WAIC difference was 13.9, with a standard deviation of 5.67. I therefore swapped out 'success' for 'internal evidence of failure' (*N*):

(Model 0.28)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = \alpha_k &+ \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FN} FN_i \\ &+ \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGN} FGN_i + \beta_{FOS} FOS_i + \beta_{FON} FON_i + \beta_{FSN} FSN_i + \\ &\beta_{FGOS} FGOS_i + \beta_{FGON} FGON_i + \beta_{FGSN} FGSN_i + \beta_{FOSN} FOSN_i + \\ &\beta_{FGOSN} FGOSN_i\end{aligned}$$

This model did improve, taking 100% of the weight compared to Model 0.26. The WAIC difference was 16.8 (SD=13.98). It was this model that was therefore graphed.

5.5: Predicting microstructure similarity scores by participant attendance to the experimental video

Model 0.29, below, used the variable 'attendance to the video' (*T*) to predict variation in microstructure similarity scores.

(Model 0.29)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i$$

The model below added the variable 'age' to Model 0.29:

(Model 0.30)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_{TG} TG_i$$

This model took all of the weight when compared with Model 0.29. The difference in WAIC scores was 15.8 (SD=9.35). Model 0.31, below, added 'female' to this model.

(Model 0.31)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_F F_i + \beta_{TG} TG_i + \beta_{TF} TF_i + \beta_{TGF} TGF_i$$

Model 0.31 made no improvement on Model 0.30, which took 0.93 of the Akaike weight. The WAIC difference between the two was 5.1 (SD=2.94). Model 0.32 swapped 'female' for 'open'.

(Model 0.32)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TGO} TGO_i$$

This model improved on Model 0.31, with a weight value of 1. The difference in WAIC scores was 21.9 (SD=11.24). Model 0.33, below, added interactions with 'social'.

(Model 0.33)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TOS} TOS_i + \beta_{TGOS} TGOS_i$$

This model was again an improvement, taking 100% of the Akaike weight to Model 0.32. The model below added 'successful'.

(Model 0.34)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TU} TU_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TGU} TGU_i + \beta_{TOS} TOS_i + \beta_{TOU} TOU_i + \beta_{TSU} TSU_i + \beta_{TGOS} TGOS_i + \beta_{TGOU} TGOU_i + \beta_{TGSU} TGSU_i + \beta_{TOSU} TOSU_i + \beta_{TGOSU} TGOSU_i$$

Model 0.34 did not improve on Model 0.33. Comparing the two, Model 0.33 took 100% of the weight, and had a WAIC score 11.9 units lower than Model 0.34 (SD=6.06). Model 0.35, below, replaced 'successful' with 'internal evidence of failure'.

(Model 0.35)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TN} TN_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TGN} TGN_i + \beta_{TOS} TOS_i + \beta_{TON} TON_i + \beta_{TSN} TSN_i + \beta_{TGOS} TGOS_i + \beta_{TGON} TGON_i + \beta_{TGSN} TGSN_i + \beta_{TOSN} TOSN_i + \beta_{TGOSN} TGOSN_i$$

This model did not improve on Model 0.33, taking 0.42 of the Akaike weight. The WAIC difference was just 0.6, with a standard deviation of 9.66, indicating considerable uncertainty about the difference between the two models. I went forward with Model 0.33 since it was able to achieve a nearly identical WAIC score to Model 0.35 despite having less information to work on. This indicates that the addition of the new 'internal evidence of failure' variable was not useful enough to overcome the overfitting risk which the greater number of parameters creates.

5.6: Predicting macrostructure similarity scores by participant attendance to the experimental video

Model 0.36, below, used the variable 'attendance to the video' (T) to predict variation in macrostructure similarity scores.

(Model 0.36)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k)=\alpha_k + \beta_T T_i$$

The model below added the variable 'age' to Model 0.36:

(Model 0.37)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k)=\alpha_k + \beta_T T_i + \beta_G G_i + \beta_{TG} TG_i$$

This new model took all of the Akaike weight. The WAIC difference was 39.3 (SD=12.78). I therefore added 'female' to Model 0.37.

(Model 0.38)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k)=\alpha_k + \beta_T T_i + \beta_G G_i + \beta_F F_i + \beta_{TG} TG_i + \beta_{TF} TF_i + \beta_{TGF} TGF_i$$

This model did not make an improvement, with only 0.11 of the weight compared to Model 0.37. The difference in WAIC was 4.1 (SD=3.35). In the next model, I substitute the variable 'female' for 'open'.

(Model 0.39)

A_i~Ordered($\mathbf{p}$)

$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TGO} TGO_i$

Model 0.39 took 0.64 of the Akaike weight compared to Model 0.37. The difference between their WAIC values was just 1.1, with a standard deviation of 5.74. I therefore continued with the Model which held the greater probability of useful predictions, by adding 'social' to Model 0.39.

(Model 0.40)

A_i~Ordered($\mathbf{p}$)

$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TOS} TOS_i + \beta_{TGOS} TGOS_i$

This was again an improved model, taking 100% of the weight to Model 0.39. The difference in WAIC scores was 29.3 (SD=12.63). I next added interactions with the variable 'successful'.

(Model 0.41)

A_i~Ordered($\mathbf{p}$)

$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_U U_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TU} TU_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TGU} TGU_i + \beta_{TOS} TOS_i + \beta_{TOU} TOU_i + \beta_{TSU} TSU_i + \beta_{TGOS} TGOS_i + \beta_{TGOU} TGOU_i + \beta_{TGSU} TGSU_i + \beta_{TOSU} TOSU_i + \beta_{TGOSU} TGOSU_i$

The model including model success again failed to improve predictions, with Model 0.41 taking none of the Akaike weight compared to Model 0.40. The difference between the two WAIC scores was 18.6 (SD=3.22). Model 0.42, below, substitutes 'successful' for participant 'internal evidence of failure'.

(Model 0.42)

A_i~Ordered($\mathbf{p}$)

$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TN} TN_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TGN} TGN_i + \beta_{TOS} TOS_i + \beta_{TON} TON_i + \beta_{TSN} TSN_i + \beta_{TGOS} TGOS_i + \beta_{TGON} TGON_i + \beta_{TGSN} TGSN_i + \beta_{TOSN} TOSN_i + \beta_{TGOSN} TGOSN_i$

This model again took 100% of the Akaike weight. The difference between the WAIC scores of Models 0.40 and 0.42 was 16.6 (SD=13.76).

Appendix 6: Further detail of results for Chapter 4

6.1: Participant age and microstructure similarity

I deal first with variation in microstructure similarity scores. Model 1 used participant age ('G') for each case in the dataset to predict variation in microstructure similarity scores across cases. Age was expected to increase microstructure similarity scores when the social model was present.

(Model 1)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GN} GN_i + \beta_{GOS} GOS_i + \beta_{GON} GON_i + \beta_{GSN} GSN_i + \beta_{GOSN} GOSN_i$$

In this model, I_i stood for the microstructure score of each individual participant. '$\mathbf{p}$' was a vector of probabilities the same length as the number of microstructure score thresholds (i.e., 15 thresholds between 16 ordinal categories), the k stood for the threshold values themselves, defined by a link to the intercept parameter α_k. Alongside G, the predictor variable representing the effect of a participant's age, O was a 1-0 predictor variable representing whether the participant built in open- rather than close-ended conditions, S was a similar 1-0 predictor representing whether the participant observed the 'social' model rather than the 'asocial' model, and N was a predictor variable which noted each participant's degree of internal evidence of failure.

Model 1 included some, but not all of the variables which were measured for the experiment. For example, the model did not include interactions with the sex of the participant or the success of the model. This was because it was found that adding these variables to Model 1 lowered its Akaike weight and raised its WAIC score. A full account of the model comparison process can be found in Appendix 5.1. Model 1 predicted that the marginal (i.e. without interaction with other variables in the model) effect of the 'age' variable on microstructure similarity scores was mostly positive. The mean effect of 'age' was estimated at 0.13 (SD=0.08). The 0.89 HPDI (Highest Posterior Density Interval) measures the range of data which accounts for 89% of the posterior probability (McElreath 2016). For the variable 'age', the 0.89 HPDI ranged from -0.01 to 0.24.

A descending slope indicates that Model 1 predicted a positive relationship between increased participant age and microstructure similarity scores. This trend seems consistent among six out of the eight interactions modelled: graphs A, C, and D in both Figures 3 and 4 (pages 58 and 59 respectively). Whether or not the children observed the social model, older children tended to use a microstructure more similar to it than younger children. The exception, consistent across both low and high internal evidence of failure, was the interaction between the asocial model and open-ended task: graph C in both Figures 3 and 4. Only in this specific condition were participants with older ages predicted to display lower microstructure similarity than younger children.

The influence of open-ended conditions (graphs B and D in Figures 3 and 4), rather than close-ended conditions (graphs A and C in Figures 3 and 4), appeared to make the effect of age on microstructure similarity less positive. When the model was asocial, across both low and high internal evidence of failure (graphs A and B in Figures 3 and 4), the change from the close- to open-ended task turned a positive relationship between age and microstructure similarity into a negative relationship. When the model was social, across both low and high internal evidence of failure (graphs C and D in Figures 3 and 4), the strength of the positive relationship between age and microstructure similarity scores was weakened by the change from a close- to an open-ended task.

The influence of the difference between low and high internal evidence of failure (Figures 3 and 4 respectively) on the effect of age on microstructure similarity scores appears to have been relatively weak. In the asocial conditions (graphs A and B in Figures 3 and 4), there seemed to have been essentially no difference between low and high participant evidence of failure. In the social conditions (graphs C and D in Figures 3 and 4), the relationship between age and microstructure similarity appears to have been slightly weaker when participants demonstrated high internal evidence of failure.

6.2: Participant age and macrostructure similarity

I then turned to the effects of the 'age' variable on macrostructure similarity scores, in which a positive relationship was again predicted when the social model was present. The macrostructure similarity data used here numbered 559 cases. The 6 cases dropped include the four also excluded from the microstructure data above, and the two builds not coded during macrostructure score data collection via Prolific. As for the microstructure data above, Model 2 (below) was the result of a process of model comparison. An account of this comparison process can be found in Appendix 5.2.

(Model 2)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{GO} GO_i + \beta_{GS} GS_i + \beta_{GN} GN_i + \beta_{GOS} GOS_i + \beta_{GON} GON_i + \beta_{GSN} GSN_i + \beta_{GOSN} GOSN_i$$

Model 2 used interactions between the same four variables ('age', 'open', 'social', and 'internal evidence of failure') to predict variation in the outcome variable: macrostructure similarity scores (A_i in the model above). The predicted marginal effect of 'age' was again largely positive. The mean effect of 'age', as estimated by the model, was 0.10 (SD=0.10; HPDI=0.89, between -0.06 and 0.26). However, to gauge Model 2's predictions of the real effect of 'age' on macrostructure similarity scores, it was necessary to simultaneously take the effects of the three other variables into account. This is possible in Figures 5 and 6 (pages 60 and 61 respectively).

The influence of the close-ended (graphs A and C) versus open-ended task (graphs B and D) appeared to have a negligible effect in Figure 5, where participants showed low internal evidence of failure. It appeared to have a greater effect, however, in Figure 6 where participants exhibited greater internal evidence of failure. The change from close- to open-ended seems to have shifted the lines up towards the top of the graph, with both the asocial (graphs A and B in Figures 5 and 6) and social model (graphs C and D in Figures 5 and 6). This indicates that the open-ended task made macrostructure similarity scores lower in these conditions.

The influence of high internal evidence of failure (Figure 6), rather than low internal evidence of failure (Figure 5), appeared also to shift the lines towards the top of the graphs, decreasing macrostructure similarity scores. However, it did not appear to change the positive relationship between age and macrostructure similarity which was found in the graphs.

6.3: Participant sex and microstructure similarity

I next turn back to microstructure similarity scores, and the role that the sex of the participant played in determining their variation. Being female, rather than male, was here predicted to increase microstructure similarity scores with the social model. The sample size again numbered 561 cases. The model to describe the role of sex, or, more specifically, the role of a participant being female rather than male, is described below. It was once again the product of a process of model comparison, an account of which can be found in Appendix 5.3.

(Model 3)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FN} FN_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGN} FGN_i + \beta_{FOS} FOS_i + \beta_{FON} FON_i + \beta_{FSN} FSN_i + \beta_{FGOS} FGOS_i + \beta_{FGON} FGON_i + \beta_{FGSN} FGSN_i + \beta_{FOSN} FOSN_i + \beta_{FGOSN} FGOSN_i$$

Model 3 used interactions between five variables to predict variation in the microstructure similarity outcome variable: ‘female’ (F), ‘age’ (G), ‘open’ (O), ‘social’ (S), and ‘internal evidence of failure’ (N).

The mean effect of a participant being female, rather than male, on microstructure similarity, as estimated by the model, was 1.40 (SD=1.45; HPDI=0.89, between -0.86 and 3.74). The effect of ‘female’ thus appears to have been more variable than that for ‘age’, with posterior probability for its effects spread widely across many possible values. To unravel Model 3’s predicted effects of being female, rather than male, on microstructure similarity scores under different conditions, it was necessary to use a set of graphs. These are shown in Figures 7, 8, 9, and 10 (pages 63 and 64 respectively).

The influence of participant age on the effect of being female on microstructure similarity appeared relatively consistent. Figures 7 and 8, with low participant age, showed the same basic trend of positive relationships between being female and microstructure similarity scores for all graphs except for graph C where there was a social model and close-ended task, which showed negative relationships. Figures 9 and 10, with high participant age, also showed broadly the same pattern as each other. In Figure 9 and 10's graphs C and D, when the model was social, there were positive relationships between being female and microstructure similarity scores. With the asocial model (graphs C and D), in both Figures 9 and 10, the positive effect of being female was less visible.

The influence of the social model (graphs A and B across Figures 7 to 10), rather than the asocial model (graphs C and D across Figures 7 to 10), on the effect of a female participant on microstructure similarity also appears to have been complex. There were examples of the social model reversing the effect of the 'female' variable on microstructure with the asocial model, such as between graphs B and D in Figure 9 with older children exhibiting lower internal evidence of failure in the open-ended task. There were also examples of the social model maintaining the effect of the female with an asocial model, such as between graphs B and D in Figure 8 with younger children exhibiting higher internal evidence of failure in the open-ended task.

The same story was true for the effect of the open-ended task (graphs B and D across Figures 7 to 10) rather than the close-ended task (graphs A and C across Figures 7 to 10). Between graphs C and D in Figure 7, with young children exhibiting low internal evidence of failure with a social model, the effect of changing open-ended into close-ended conditions appears to reverse the effect of being female on microstructure similarity. However, between graphs A and B also in Figure 7, when the model was asocial, the effect of close- versus open-ended conditions did not seem to change the positive relationship between being female and higher microstructure similarity scores.

The influence of internal evidence of failure also appeared to have complex influences on the effect of being female on microstructure similarity scores. Between Figures 7 and 9, where participants displayed low internal evidence of failure, there were reversals in the direction of the relationship between being female and microstructure similarity in three out of the four graphs. Between Figures 8 and 10, with high internal evidence of failure, there were two reversals in the relationships between being female and microstructure similarity scores.

6.4: Participant sex and macrostructure similarity

I now consider the effect of being female, the same variable, on variation in macrostructure similarity scores. The effect of female was again predicted to increase the similarity of participants' builds to the social model when the social model was present. Model 4, below, used the same five predictor variables as above to model variation in macrostructure similarity scores across the 559 cases. See Appendix 5.4 for an account of the model comparison process.

(Model 4)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_F F_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{FG} FG_i + \beta_{FO} FO_i + \beta_{FS} FS_i + \beta_{FN} FN_i + \beta_{FGO} FGO_i + \beta_{FGS} FGS_i + \beta_{FGN} FGN_i + \beta_{FOS} FOS_i + \beta_{FON} FON_i + \beta_{FSN} FSN_i + \beta_{FGOS} FGOS_i + \beta_{FGON} FGON_i + \beta_{FGSN} FGSN_i + \beta_{FOSN} FOSN_i + \beta_{FGOSN} FGOSN_i$$

The mean effect of 'female' was more firmly positive than in Model 3 for microstructure outcome variable, at 4.51 (SD=1.78; HPDI=0.89, between 1.69 and 7.39). Figures 11 to 14 (see pages 65, 66, and 67) illustrate the effects of being female in interaction with the other variables.

The social model did not appear to have a totally consistent influence on the effect of being female on macrostructure similarity scores. In Figure 11, with younger children exhibiting lower internal evidence of failure, quite strong effects of being female on macrostructure similarity with an asocial model (graphs A and B) were made much weaker with a social model (graphs C and D). However, in Figure 14, with older children exhibiting higher internal evidence of failure, the change of the asocial (graphs A and B) to the social model (graphs C and D) appears to have inverted the relationship between being female and macrostructure similarity scores.

The influence of an open-ended task (graphs B and D in Figures 11 through 14), rather than a close-ended task (graphs A and C in Figures 11 through 14), was variable. In five out of eight conditions, the change from the close-ended to the open-ended task made the effect of being female on macrostructure similarity at least as positive as in the close-ended task (in four of these cases, the effect of being female became more positive). This was true, for example, in Figure 13 with older children demonstrating lower internal evidence of failure. Between graphs A and B, with the asocial model, the influence of the open-ended task was to reverse the negative relationship between being female and macrostructure similarity found with the close-ended task. Between graphs C and D, with a social model, the influence of the open-ended task was to turn a neutral relationship between being female and macrostructure similarity into a positive one. However, there were three more conditions in which the open-ended condition was associated with either a negative or neutral relationship between being female and macrostructure similarity scores. Between graphs A and B in Figure 14, with older children demonstrating higher internal evidence of failure and an asocial model, for example, the influence of open-ended conditions appeared to have been to conserve the slightly negative relationship between being female and macrostructure similarity scores in the close-ended condition. This indicates that the effect of the open-ended task, versus close-ended task, was itself dependent on the influence of other variables, particularly the social versus asocial model.

In the set of graphs between Figures 11 and 14, the influence of age did not appear to have been a large factor in the effect of being female on macrostructure similarity scores. Within the younger and older age groups there was a large degree of variation, across both Figures 11 and 12 (younger) and both Figures 13 and 14 (older). And between Figures 11 and 13, and 12 and 14 respectively, there was a surprising degree of similarity. This suggests that variation in internal evidence of failure may have been a strong predictor of the effect of being female on macrostructure similarity scores. However, even here, within conditions of either low or high internal evidence of failure, there was a large degree of variation. For example, between Figures 12 and 14, where participants showed high internal evidence of failure, the effects of 'female' on macrostructure similarity were: inverted from positive to negative in graph A, kept negative in graph B, inverted from negative to positive in graph C, and kept positive in graph D. The influence of internal evidence of failure thus also appeared to have been dependent on the influences of other variables.

6.5: Participant attendance to the video and microstructure similarity

The final variable to be tested in Chapter 4 was participants' attendance to the video. Higher attendance to the video was predicted to increase microstructure similarity scores with the social model. Model 5, below, used four predictor variables to model variation in microstructure similarity across the 561 cases: 'attendance to the video' (T), 'age' (G), 'open' (O), and 'social' (S).

(Model 5)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TOS} TOS_i + \beta_{TGOS} TGOS_i$$

Model 5 did not include interactions with the 'internal evidence of failure' variable, since adding it was not found to increase model predictions. An account of these model comparisons can be found in Appendix 5.5. Model 5's estimate for the mean effect of the attendance score variable (without interactions with other variables) was 0.04 (SD=0.11; HPDI=0.89, between -0.13 and 0.21). Figures 15 and 16 (pages 68 and 69 respectively), on the other hand, plot Model 5's predicted real effects of turning participant attendance to the video from low to high, when its effect was dependent on the influence of the three other variables in the model: the close- versus open-ended task, the asocial versus social model, and low versus high participant age.

The influence of the social, rather than asocial, model appeared to have little change amongst younger children (Figure 15). In the close-ended task, the social model (in graph C) changed the weakly positive effect of attendance scores on microstructure similarity with the asocial model (graph A) into a weakly negative effect. With the open-ended task, the social model (graph D) merely conserved the already weakly negative relationship between attendance scores and microstructure similarity visible with the asocial model (graph B). In older children (Figure 16), however, the effect of the social model appeared to reduce the negative effect of attendance to the video on microstructure similarity scores. With the close-ended task, the social model (in graph C) changed the asocial model's neutral relationship between attendance and microstructure similarity (graph A) into a positive relationship, and with the open-ended task, the social model (in graph D) appeared to nullify the negative relationship between attendance and microstructure similarity found with the asocial model (graph B).

The influence of open-ended, rather than close-ended, conditions appears to have been to make the effect of attendance to the video on microstructure similarity scores more negative, or at least less positive. In Figure 15 (with younger children), both open-ended conditions (graphs B and D) exhibited weakly negative relationships between attendance to the video scores and microstructure similarity scores, while the close-ended conditions showed either an even more weakly negative relationship (with the social model, graph C) or a weakly positive relationship (with the asocial model, graph A). In Figure 16 (amongst older children), with an asocial model, while the close-ended condition (graph A) showed a neutral relationship, the open-ended condition (graph B) showed a negative relationship. With a social model, the close-ended task (graph C) showed a relatively strong positive relationship between attendance to the video and microstructure similarity, which was nullified in the open-ended task (graph D).

The influence of participant age appeared flexible. Figure 16, with the older children, presented the two graphs with the strongest relationships between attendance scores and microstructure similarity scores, as well as the two graphs with the weakest relationships between the predictor and outcome variables. Older age seemed to enhance the negative effect of attendance to the video with an asocial model and open-ended task, as well as the positive effect of attendance to the video with a social model and close-ended task. Older age then seemed to nullify the effect of attendance scores on microstructure similarity with an asocial model and close-ended task, and with a social model and open-ended task.

6.6: Participant attendance to the video and macrostructure similarity

Finally, for Chapter 4, I considered the effects of participants' attendance to the video on macrostructure similarity scores. The effect of attendance to the video was again predicted to be positive when the social model was present. The sample size was 561. The model, below, used interactions between five variables: 'attendance to the video' (T), 'age' (G), 'open' (O), 'social' (S), and 'internal evidence of failure' (N). See Appendix 5.6 for an account of the comparisons to arrive at this model.

(Model 6)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = \alpha_k &+ \beta_T T_i + \beta_G G_i + \beta_O O_i + \beta_S S_i + \beta_N N_i + \beta_{TG} TG_i + \beta_{TO} TO_i + \beta_{TS} TS_i + \\ &\beta_{TN} TN_i + \beta_{TGO} TGO_i + \beta_{TGS} TGS_i + \beta_{TGN} TGN_i + \beta_{TOS} TOS_i + \beta_{TON} TON_i + \\ &\beta_{TSN} TSN_i + \beta_{TGOS} TGOS_i + \beta_{TGON} TGON_i + \beta_{TGSN} TGSN_i + \beta_{TOSN} TOSN_i + \\ &\beta_{TGOSN} TGOSN_i\end{aligned}$$

Model 6's estimated mean effect of the attendance to the video variable was positive, at 0.37 (SD=0.22; HPDI=0.89, between 0.04 and 0.74). The predicted effects of participant attendance to the video on macrostructure similarity scores, when interacting with the four other variables in Model 6, are shown in Figures 17/23, 18/24, 19/25, and 20/26 (see pages 70 to 72).

The influence of internal evidence of failure again appears to have been relatively minor. The direction of the effects of 'attendance to the video' on macrostructure similarity scores were conserved in arguably only three out of the eight comparisons between low and high internal evidence of failure (between Figure 17 and 18's graph 'A's, Figure 19 and 20's graph 'A's, and Figure 19 and 20's graph 'B's), with three clear reversals of the effects of attendance scores on macrostructure similarity scores (between Figure 17 and 18's graph

'B's, 'C's, and 'D's). The pattern underlying these visible differences did not seem to have been the product consistent effects of the other predictor variables. This indicates that the role internal evidence of failure took in determining the effect of attendance scores on macrostructure similarity was itself dependent on the interaction of various other predictor variables.

The influence of the social, rather than asocial, model appears to have been different where participant internal evidence of failure was low and high. Where internal evidence of failure was 'low', across open- and close-ended conditions and low and high participant ages, the graphs with the social model (graphs C and D in Figures 17 and 19) show less influence of attendance scores on macrostructure similarity than with the asocial model (graphs A and B in the same two Figures). However, where internal evidence of failure was high, the effects of 'attendance score' appear to have been increased with a social model (graphs C and D in Figures 18 and 20) rather than an asocial one (graphs A and D in the same two Figures).

The influence of the open-ended task, rather than the close-ended task, appears to have had little impact on the effect of attendance to the video scores on macrostructure similarity. Between all of the eight comparisons between close- and open-ended conditions (between graphs A and B and between graphs C and D in Figures 17, 18, 19, and 20), the relationship between attendance score and macrostructure similarity appears very similar. Perhaps the one exception to this was in Figure 18, with younger children exhibiting higher internal evidence of failure and an asocial model (graph A), where a positive relationship between attendance to the video and macrostructure similarity was made negative by changing the task from open- to close-ended.

The influence of participant age also seems to have been a factor here, though more so where internal evidence of failure was low, and more so with the asocial model. Between Figures 17 and 19 (with low internal evidence of failure and, respectively, low and high participant age), with the asocial model (i.e., graphs A and B), the direction of the effect of 'attendance to the video' appears to reverse.

With the social model (graphs C and D), however, lack of a directional influence of attendance scores on macrostructure similarity was maintained across low and high participant age. The greatest difference between low and high ages seems to have been with the asocial model and close-ended task (graph A), while conditions with the asocial model and open-ended task (graph B) and the two conditions with the social model (graphs C and D) show less change between low and high participant age. Between Figures 18 and 20 (with high internal evidence of failure and, respectively, low and high participant age), the positive effects of 'attendance to the video' were maintained in graphs C and D (with the social model). Meanwhile, while Figure 18 shares graph B's negative effect of 'attendance to the video' with Figure 20, the effects of 'attendance to the video' on macrostructure similarity scores in graph A were reversed between Figures 18 and 20.

Appendix 7: Model descriptions and comparisons for Chapter 5

7.1: Hypothesis 1

The description of the first model, using only 'close', here labelled C, as a predictor of variation is:

(Model 1.1)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_C C_i$$

In the next model, I use the 'close' variable to predict variation in macrostructure similarity scores but make its effect dependent on whether the participant was building under social or asocial conditions (the variable S below).

(Model 1.2)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_{CS} CS_i$$

Comparing these two models, Model 1.2 captures all of the Akaike weight, leaving Model 1.1 with none. The difference between their WAIC scores was 17.7, with a standard deviation of 9.37 indicating a substantial difference in the predictive power of the two models. Next, I added a variable which recognises a difference between the experimental model building successfully or unsuccessfully ('successful', variable U).

(Model 1.3)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_U U_i + \beta_{CS} CS_i + \beta_{CU} CU_i + \beta_{CSU} CSU_i$$

In comparison with Model 1.2, Model 1.3 receives 0.14 of the Akaike weight, and its WAIC score was 3.6 units higher (SD=3.35) than Model 1.2's. This indicates that the addition of the 'successful' variable causes overfitting of the model, which damages its ability to make predictions about future data. This would therefore suggest that the success of the model was not a particularly useful piece of information for making predictions about, and therefore understanding the variation in, macrostructure scores here. This was further supported by a model similar to Model 1.3, but which excludes the three-way interaction term between 'close', 'social', and 'successful'. This Model 1.4 gained only 0.21 of the weight compared to Model 1.2, with a difference of 2.6 in WAIC values

(SD=2.54). Model 1.5, below, thus swapped out 'successful' for 'internal evidence of failure' (N), to test whether this variable was more useful in understanding the present variation.

(Model 1.5)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_N N_i + \beta_{CS} CS_i + \beta_{CN} CN_i + \beta_{CSN} CSN_i$$

This time the addition of a variable was decisively positive. The new model, Model 1.5, gained 100% of the Akaike weight in comparison to Model 1.2. Its WAIC value was lower than Model 1.2's by 40.5 units (SD=14.73). However, the success of the participant may arguably have been related to the success of the model. So below, Model 1.6 tests hierarchical interactions between all four of the variables considered so far.

(Model 1.6)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_{CS} C_i S_i + \beta_{CU} C_i U_i + \beta_{CN} C_i N_i + \beta_{CSU} C_i S_i U_i + \beta_{CUN} C_i U_i N_i + \beta_{CSN} C_i S_i N_i + \beta_{CSUN} C_i S_i U_i N$$

This model did appear to have been an improvement on Model 1.4, a comparison with which showing Model 1.6 to have a lower WAIC value by 38.4 (SD=15.95) and to take 100% of the weight. Yet Model 1.6 also appears to have made poorer out-of-sample predictions than Model 1.5. Compared with Model 1.5, Model 1.6 took only 0.09 of the Akaike weight. The difference between the two WAIC values was 4.7, with a standard deviation of 5.42. This large standard deviation indicates there was uncertainty about which model made the better predictions. Nevertheless, the greater probability was that Model 1.6 made worse predictions than Model 1.5, whilst also being a more complicated model. It was therefore with Model 1.5 that I proceeded to test the other predictor variables: age, sex, and attendance to the experimental video. In Model 1.7, below, I added a predictor variable which took account of the children's ages (G), which vary from 5 to 11 years old, to Model 1.5. Its description is:

(Model 1.7)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_C C_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_{CS} CS_i + \beta_{CN} CN_i + \beta_{CG} CG_i + \beta_{CSN} CSN_i + \beta_{CSG} CSG_i + \beta_{CNG} CNG_i + \beta_{CSNG} CSNG_i$$

Model 1.7 took 100% of the Akaike weight relative to Model 1.5, with a WAIC lower WAIC score by 34.3 (SD=12.43). It also took 100% of the Akaike weight relative to Model 1.6, with a larger 77.4 difference in WAIC scores (SD=18.25). Participant age thus seems to have been an important contributing factor in assessing the influence of close-ended conditions on macrostructure similarity scores. In the next model, I tried adding the variable 'female' to Model 1.7, and incorporating it into the interactions between other variables.

(Model 1.8)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_C C_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_F F_i + \beta_{CS} CS_i + \beta_{CN} CN_i + \beta_{CG} CG_i + \beta_{CF} CF_i + \beta_{CSN} CSN_i + \beta_{CSG} CSG_i + \beta_{CSF} CSF_i + \beta_{CNG} CNG_i + \beta_{CNF} CNF_i + \beta_{CGF} CGF_i + \beta_{CSNG} CSNG_i + \beta_{CSNF} CSNF_i + \beta_{CSFG} CSFG_i + \beta_{CSNFG} CSNFG_i$$

The addition of the variable 'female', and interactions with it, reduces the ability of the model to predict future data. Compared to Model 1.7, Model 1.8 took none of the Akaike weight. The WAIC difference between the two was 13 (SD=6.76). However, it may be argued that this model failed due to too large a number of parameters, rather than the specific effects of including 'female' as a variable. To test this, I replaced the 'age' variable in Model 1.7 with the 'female' variable and compared the new model (Model 1.9) to the original Model 1.7. The comparison reveals that the original Model 1.7 took all of the weight. The difference in WAIC scores between them was 38.1, with a standard deviation of 15.47. This supports the claim that introducing the variable 'age' provides useful information for examining the effect of social conditions on macrostructure similarity scores, whilst introducing the variable 'female' did not. Model 1.10 below was the same as Model 1.8, except used the variable 'attendance to the video' (T) instead of 'female'.

(Model 1.10)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_C C_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_T T_i + \beta_{CS} CS_i + \beta_{CN} CN_i + \beta_{CG} CG_i + \beta_{CT} CT_i + \beta_{CSN} CSN_i + \beta_{CSG} CSG_i + \beta_{CST} CST_i + \beta_{CNG} CNG_i + \beta_{CNT} CNT_i + \beta_{CGT} CGT_i + \beta_{CSNG} CSNG_i + \beta_{CSNT} CSNT_i + \beta_{CSTG} CSTG_i + \beta_{CSNTG} CSNTG_i$$

Like Model 1.8, Model 1.10 had an Akaike weight of zero relative to Model 1.7. The WAIC difference between the two was 14.9 (SD=4.57). And like Model 1.8, the same argument could be made: that the model failed due to overfitting with too great a number of parameters. Thus I also swapped out the ‘age’ variable from Model 1.7 and replaced it with ‘attendance to the video’ to create a new model (Model 1.11), as before. Here the difference between either including ‘attendance to the video’ or including ‘age’ was starker: Model 1.7 took 100% of the weight, and the WAIC difference between the two models was 42.7 (SD=14.35). These comparisons indicated that the ‘age’ variable was useful for making predictions about the effect of close-ended conditions on macrostructure similarity scores, that adding either ‘female’ or ‘attendance to the video’ to the model results in worse predictions for the data, and that replacing ‘age’ with either ‘female’ or ‘attendance to the video’ also results in worse predictions for the data. The model that I therefore continued my analyses with was Model 1.7, since participants’ sex and attendance to the video appears to have been relatively less useful, and therefore relatively less interesting, in considering the impact of close-ended conditions. In Model 1.12, described below, I remove ‘close’ as the main predictor variable from Model 1.7.

(Model 1.12)

$A_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{NG} NG_i + \beta_{SNG} SNG_i$

Surprisingly, a comparison between Models 1.12 and 1.7 would suggest that the close-ended task was not that an important predictor of variation in macrostructure similarity scores. Model 1.12 gained 100% of the Akaike weight and a WAIC score which was 12.3 units lower than that of Model 1.7, the standard deviation of which was 7.32. This indicates that adding interactions with the ‘close’ variable decreases the ability of the model to make predictions about new data from the same experiment. This in turn indicates that the effect of the ‘close’ variable may not have been very significant.

7.2: Hypothesis two

The first model used only ‘social’ as a predictor of variation in microstructure similarity scores. Its description is:

(Model 2.1)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k)=\alpha_k + \beta_S S_i$

However, Model 2.1 did not take account of other causes of variation in the microstructure data. The effect of social models may have been different depending on other elements of the experiment, such as the success of the model in question. The model below thus added the variable 'success' to Model 2.1.

(Model 2.2)

$I_i\sim\text{Ordered}(\mathbf{p})$

$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_U U_i + \beta_{SU} SU_i$

A comparison between the two models reveals that Model 2.2 took 0.8 of the Akaike weight to Model 2.1's 0.2. The difference in WAIC scores was 2.7, but with a standard deviation of 5.54. Thus both variables appear important in predicting the variation of microstructure similarity scores in this condition. Model 2.3 below expands on this by adding an 'internal failure' variable (N) to Model 2.2, and interactions with 'social' and 'success'.

(Model 2.3)

$I_i\sim\text{Ordered}(\mathbf{p})$

$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_{SU} SU_i + \beta_{SN} SN_i + \beta_{SUN} SUN_i$

Compared to Model 2.2, Model 2.3 took all of the Akaike weight. The difference in WAIC scores between the two was 12.2 (SD=8.33). This indicates that 'internal failure' provides a parameter useful for making predictions from the data, which outweighs its effects on the possibility of overfitting the model to the data (McElreath 2016:166). I therefore proceeded with Model 2.3. There may have been other sources of variation which influence the effect of social models on microstructure similarity scores. One of these was participant age, so Model 2.4 below added an interaction between 'social' (S) and 'age' (G).

(Model 2.4)

$I_i\sim\text{Ordered}(\mathbf{p})$

$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_{SU} SU_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{SUN} SUN_i$

The comparison between Models H2.3 and H2.4 reveals that H2.4 took all of the weight, with the difference in WAIC scores 30.3 (SD=13.16). Age therefore seems an important contributing factor in how a social model impacts microstructure similarity scores. Model 2.5 below thus added the variable ‘female’ to see whether differences in sex have a similarly important effect.

(Model 2.5)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_F F_i + \beta_{SU} SU_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{SF} SF_i + \beta_{SUN} SUN_i$$

Compared to Model 2.4, Model 2.5 had a lower Akaike weight, taking 0.15 to Model 2.4’s 0.85. Model 2.5 also had a higher WAIC value than Model 2.4, the difference being 3.4 units (SD=2.54). This suggests that the variable ‘female’ causes more overfitting than its contribution to prediction was worth, which indicates that the difference between male’s and female’s scores did not add a great deal of important information for understanding the influence of social models on microstructure similarity score variation. The same finding was found when ‘female’ was replaced by ‘attendance to the video’ (compared to Model 2.4, a weight of 0.13 and WAIC difference of 3.8, SD=1.09), indicating that this measure was similarly less significant for understanding the impact of a social model. It could be argued, however, that Models 2.5 and 2.6 failed due to overfitting too many parameters. Therefore I also calculated comparisons between Model 2.4 and the same model but where ‘age’ was replaced first by ‘female’ (Model 2.7) and then by ‘attendance to the video’ (Model 2.8). Neither of these models take any weight in a comparison with Model 2.4. The difference between the WAIC scores of Model 2.4 and Model 2.7 was 32.9 (SD=13.51), while the difference with Model 2.8 was 33.8 (SD=13.18). Therefore it was Model 2.4 that I proceeded with, since it was this that had the best comparison scores of any of the models tested for this data. This was supported by a comparison of Model 2.4 to a similar model but without a variable for ‘social’ (Model 2.9), and a model in which all of the parameters tested so far interact in a hierarchical pattern with the ‘social’ variable (Model 2.10). Model 2.9 gained no weight relative to Model 2.4 (with a WAIC difference of 36.5, SD=12.72), as did Model 2.10 (with a WAIC difference of 44.2, SD=11.81). However, to fully understand the role that participant age plays in the effect of social models on microstructure similarity scores, I need to more fully integrate it into interactions with other variables. When this was done, the model looks like this:

(Model 2.11)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_{SU} SU_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{SNG} SUN_i + \beta_{SUN} SUG_i + \beta_{SUG} SNG_i + \beta_{SUNG} SUNG_i$$

7.3: Hypothesis 3

The first model fitted predicts variation in microstructure similarity scores purely with the success of the model (U).

(Model 3.1)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_U U_i$$

The model below added the variable for internal evidence (N) of failure to Model 3.1.

(Model 3.2)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_U U_i + \beta_N N_i + \beta_{UN} UN_i$$

Model 3.2 gained 0.99 of the Akaike weight relative to Model 3.1, with a difference in WAIC scores of 9.1 (SD=6.43), indicating a better ability to predict future data from the same experiment. Model 3.3 below added interactions with the variable 'age' (G) to this model.

(Model 3.3)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UNG} UNG_i$$

This model again improved on Model 3.2's weight value. In a comparison between Model 3.2 and Model 3.3, Model 3.3 took 100% of the weight. Model 3.3 had a lower WAIC score than Model 3.2 by 21.3 units (SD=11.92). I then tried adding the other predictor variables to Model 3.3: 'female' and 'video attendance score'. The description below shows the addition of 'female' (F).

(Model 3.4)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k)=\alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_F F_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UF} UF_i + \beta_{UNG} UNG_i + \beta_{UNF} UNF_i + \beta_{UGF} UGF_i + \beta_{UNGF} UNGF_i$$

However, each of the new predictor variables ('female' and 'video attendance score') lowered the Akaike weight, indicating that they did more harm than good in predicting variation in microstructure similarity. These scores can be seen in first two rows of Table 6. It could be argued, however, that the models failed because they contained too many parameters, rather than the usefulness of the variables. The bottom two rows of Table 6 thus show the differences in Akaike weight and WAIC scores between Model 3.3 and two models when the 'age' variable in Model 3.3 was swapped out for either 'female' or 'attendance to video score'. When the addition of 'female' was found to lower the weight value, it was not kept for the subsequent model. Therefore each row of Table 6 represents an independent test of the variable in interaction with the other variables of Model 3.3. Since it appears that inclusion of either 'female' or 'video attendance score', the model comparison implies that neither of these variables were useful predictors of variation in microstructure similarity scores in these conditions. In other words, they were not particularly important factors in explaining the observed variation. Instead, these model comparisons indicated that the key factors to consider are: model success (external evidence of failure), participant success (internal evidence of failure), and participant age. This was further supported by model comparisons. I undertook comparisons between Model 3.3 and a model (Model 3.8) including all of the variables in the various models above in hierarchical levels of interaction with one another, and between Model 3.3 and a model (Model 3.9) including just interactions between participant success (N) and participant age (G). This latter model's description is:

(Model 3.9)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_G G_i + \beta_{NG} N G_i$$

Table 6

Results of model comparisons between Model 3.3 and Models 3.4, 3.5, 3.6, and 3.7.

Model number	Variable tested	Weight relative to Model 3.3	Difference in WAIC to Model 3.3	Standard deviation of the WAIC difference
Model 3.4	'Female'	0	10.9	2.85
Model 3.5	'Video attendance score'	0.1	9.5	3.2
Model 3.6	'Female'	0	26.3	12.4
Model 3.7	'Video attendance score'	0	26.9	12.03

Note. Models 3.4 and 3.5 each added an additional variable to Model 3.3. The results show that both of these additions reduced the Akaike weight value and WAIC score, indicating that these two variables did not make better models of microstructure similarity score variation when they were included than when they were not.

While Model 3.8 had a weight of zero, Model 3.9 gained a weight of 0.43, and a WAIC value of only 0.5 below Model 3.3's (SD=5.87). I therefore proceeded with Model 3.3, the model which had the lowest WAIC value and highest Akaike weight of any of the models tested. However the closeness of Model 3.9 may indicate that the effect of successful models on participants' microstructure similarity scores may have been relatively weak, when taking other relevant factors into account.

In testing the hypothesis that the success of the model was a positive predictor of microstructure similarity score variation, it was necessary to have a control group to test whether higher similarity scores were the product of children's copying rather than other factors. This meant including the 'asocial' conditions into the dataset, and introducing a new variable to indicate the 'social' versus 'asocial' status of cases. The data for the hypothesis therefore now numbered 273 cases. Model 3.3, with the addition of a 'social' variable, now looks like:
(Model 3.10)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_S S_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{US} US_i + \beta_{UNG} UNG_i + \beta_{UNS} UNS_i + \beta_{UGS} UGS_i + \beta_{UNGS} UNGS_i$$

In a comparison between Models 3.10 and 3.3, on the new larger dataset, Model 3.10 took all of the Akaike weight. The difference between their two WAIC scores was 27, with a standard deviation of 12.05. This indicates that despite the increased number of parameters, with its risk of overfitting, the larger model still made better predictions about the data, indicating that the effect of model success was dependent on whether the model was social or asocial. This constitutes preliminary evidence supporting the idea that it was copying by which participants have increased microstructure similarity scores when observing successful social models.

7.4: Hypothesis four

Hypothesis 4 states that across close-ended social model conditions, the success of the model will not be a good predictor of variation in macrostructure similarity scores. This was because macrostructure diversity was constrained by the close-ended setup. The sample size of the data for the models below also numbered 144 participants. This number was reached by subtracting the two builds for which macrostructure were not coded, the cases for which other data were absent, and those participants who built under asocial or open-ended conditions. The model described below used only the success of the model to predict macrostructure score variation.

(Model 4.1)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i$$

I then ran a model which introduced evidence of internal (i.e., participant) failure as another source of variation in macrostructure similarity scores, which may interact with the effect of model success.

(Model 4.2)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_{UN} UN_i$$

This model improved on Model 4.1's WAIC score, at 6.8 units down (SD=8.11). The comparison between the two reveals that Model 4.2 took 0.97 of the weight. This indicates that the interaction of 'successful' with 'internal evidence of failure' was useful for understanding the observable variation in macrostructure similarity scores. I now test whether the same was true when interaction with 'age'.

(Model 4.3)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UNG} UNG_i$$

Model 4.3 was a clear improvement on Model 4.2. It took all of the Akaike weight, and had a WAIC score 15.8 units lower than Model 4.2 (SD=11.09). Participant age again seems to have been a key auxiliary variable in the effect of the hypothesised predictor variable. The model below added 'female' to Model 4.3.

(Model 4.4)

$$A_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_F F_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UF} UF_i + \beta_{UNG} UNG_i + \beta_{UNF} UNF_i + \beta_{UGF} UGF_i + \beta_{UNGF} UNGF_i$$

Compared to Model 4.3, this model had a weight of zero. Model 4.4's WAIC score was 12.2 units above that of Model 4.3's, with the standard deviation of this difference being 3.93. The model below (Model 4.5) replaced the 'female' variable in Model 4.4 with a variable for 'attendance to the video'.

(Model 4.5)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_T T_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{UF} UF_i + \beta_{UNG} UNG_i + \beta_{UNF} UNF_i + \beta_{UGF} UGF_i + \beta_{UNGF} UNGF_i$$

In comparison to Model 4.3, this model also gained an Akaike weight of zero. Model 4.5's WAIC score was 11.8 units above that of Model 4.3's, with the standard deviation of this difference 4.64. However, for both of these models, as for the hypotheses above, this difference may have been due to the number of parameters rather than the nature of the variables they represent. Thus I computed two models in which the 'age' variable of Model 4.3 was replaced by 'female' and 'attendance to the video', respectively named Models 4.6 and 4.7. Both of these new models, 4.6 and 4.7, resulted in Akaike weights of zero in comparison to Model 4.3. The difference in WAIC values between Models 4.3 and 4.6 was 21.5 (SD=11.49), while between Models 4.3 and 4.7 it was 21.8 (SD=11.16). These results show that neither the 'female' or 'attendance to the video' variables were either more useful than 'age' for estimating the effect of successful models on macrostructure similarity scores, and that neither 'female' nor 'attendance to the video' were useful for estimating the effect of successful models on macrostructure similarity scores interaction with 'age'. Therefore, further analyses proceeded with Model 4.3, the model with the highest weight and lowest WAIC scores of all models tested for this hypothesis thus far.

However, in testing the hypothesis that the success of the model was not a good predictor of macrostructure similarity score variation, it was necessary to have a control group to test whether the higher similarity scores were the product of children's copying rather than other factors. This meant including the 'asocial' conditions into the dataset, and introducing a new variable to indicate the 'social' versus 'asocial' status of cases. The data for the hypothesis therefore now numbered 273 cases. Model 4.3, with the addition of a 'social' variable, now looks like:

(Model 4.8)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_N N_i + \beta_G G_i + \beta_S S_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{US} US_i + \beta_{UNG} UNG_i + \beta_{UNS} UNS_i + \beta_{UGS} UGS_i + \beta_{UNGS} UNGS_i$$

Compared with Model 4.3, with both having access to the new larger dataset, Model 4.8 took 0.99 of the Akaike weight, with the WAIC value difference between the two at 9.6 units (SD=11.68). The overlap of the standard deviation in the difference between WAIC scores indicates that Model 4.8 may have overfitted the data slightly. However, the order of the WAIC scores and the degree of difference in the Akaike weight measure, paired with the fact that Model 4.8 did not lose any of the information which Model 4.3 has, means that Model 4.8 appears to have been most useful for the purpose of testing the current hypothesis.

Appendix 8: Further analyses of data for Chapter 5

8.1: Hypothesis 1

The effect of older (Figure 23, page 77), rather than younger (Figure 22, page 76) children generally appeared to enhance the positive effect of close-ended conditions with a social model (graphs C and D), while dampening the positive effect of 'close' with an asocial model (graphs A and B). Accordingly, the influence of younger children (Figure 22) appeared to enhance the positive effect of 'close' in asocial conditions (graphs A and B) whilst dampening its positive effect in social conditions (graphs C and D). The effect of the social, rather than the asocial, model thus appeared to have been dependent on the age of the participant. In younger children (Figure 22), there was a greater positive effect of 'close' on macrostructure similarity when the model was asocial (graphs A and B) rather than social (graphs C and D). However, for older children (Figure 23), while there appeared to have been little increase in the strength of the positive effect of 'close' on macrostructure similarity when internal evidence of failure was low (graphs A and C), when internal evidence of failure was high (graphs B and D) the positive effect of 'close' on macrostructure similarity was strengthened. Similarly, the influence of higher internal evidence of failure (graphs B and D) appeared to have been associated with a greater effect of close-ended conditions on macrostructure similarity scores in older children (Figure 23) but not younger (Figure 22).

8.2: Hypothesis 2

The effect of model success appeared to increase the positive effect of the social model. In younger children (Figure 25, page 81), this meant that the successful model (graphs C and D, rather than the unsuccessful model in graphs A and B) made the difference between a social model having no discernable or even a negative effect on microstructure similarity, and a social model having a clearly positive effect on microstructure similarity. In older children (Figure 26, page 82), the effect of the successful model (graphs C and D) rather than the unsuccessful model (graphs A and B) made the positive effect of a social model clearer, which held true across both low (graphs A and C) and high internal evidence of failure (graphs B and D).

The effect of high internal evidence of failure (graphs B and D) appeared to make the effect of a social model on microstructure similarity less clearly positive in younger children (Figure 25). In younger children with an unsuccessful model (graphs A and B), the presence of high internal evidence of failure (graph B) even appeared to change the effect of a social model from a negligible effect in either direction into a notably negative influence on microstructure similarity scores. These effects of internal evidence of failure did not seem present, however, amongst older children (graphs A to D in Figure 26).

Whether with successful (graphs C and D) or unsuccessful models (graphs A and B), and whether with low (graphs A and C) or high internal evidence of failure (graphs B and D), older children (in Figure 26) were predicted to have a stronger positive effect of a social model on microstructure similarity than younger children (in Figure 25). For children building in conditions with an unsuccessful model and experiencing high internal evidence of failure (graph B), the direction of the influence of a social model on microstructure similarity was reversed between younger and older children (i.e., between Figures 25 and 26).

8.3: Hypothesis 3

High rather than low internal evidence of failure (graphs B and D) appeared to reduce the influence of model success on microstructure similarity seems true in every condition (Figures 28 and 29, pages 86 and 87 respectively), including in older children with an asocial model (graph B, Figure 29). When these children exhibited low internal evidence of failure, there was a reliable effect of model success to decrease macrostructure similarity. This became messier and less clear when the older children instead exhibited high internal evidence of failure (Figure 29's graph D).

The influence of the higher children's age on the effect of successful models seems to have been greater for asocial than social models. Graphs A and B of both Figures 28 and 29 appear relatively similar. However, the slopes of graphs C and D in Figure 28 were inverted in Figure 29. Whereas for older children (Figure 29) asocial model success caused reduced microstructure similarity (between graphs A and B), in younger children (Figure 28) it caused increased microstructure similarity. This was clearer with low internal evidence of failure (in graph A), but the same trend, albeit in a much weaker form, was visible for high internal evidence of failure as well (in graph B).

8.4: Hypothesis 4

The influence of high internal evidence of failure (graphs B and D) here seemed to make the effect of the successful model on macrostructure similarity less positive. In the case of the asocial model, in both younger and older children (Figures 31 and 32, on pages 91 and 92 respectively), high internal evidence of failure (graph B) made the effect (on macrostructure similarity scores) of turning an unsuccessful model into a successful model negative. In the case of the social model, in both younger and older children (Figures 31 and 32), the effect of high internal evidence of failure (graph D, relative to graph C's low internal evidence of failure) was to lessen the positive effect of a successful model on macrostructure similarity.

In asocial model conditions with low internal evidence of failure (graph A), neither younger children (in Figure 31) nor older (in Figure 32) showed a clearly positive effect of model success. Furthermore, both younger and older children showed a negative relationship between asocial model success and macrostructure similarity when internal evidence of failure was high (graph B). When the change in model success occurred with a social model, there were much clearer differences between younger and older children. In conditions of both low and high internal evidence of failure (graphs C and D), older children showed a more clearly positive effect of model success on macrostructure similarity.

Appendix 9: Model descriptions and comparisons for Chapter 6

9.1: Hypothesis one

The first model fitted used just model success as a predictor:

(Model 1.1)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_U U_i$

To this model, I introduced another variable for 'social' (S), which discriminated between participants observing a social model (permitting the possibility to copy) and those observing an asocial (i.e. irrelevant) model.

(Model 1.2)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_{US} US_i$

Adding this interaction effect increases the predictive power of the model for these data. In a comparison of Models 1.1 and 1.2, Model 1.2 took 100% of the Akaike weight, and had a lower WAIC score by 13.2 units (SD=8.29). The model below added an interaction of these two variables with a new variable, 'internal evidence of failure' (N):

(Model 1.3)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{USN} USN_i$

This was again an improvement on the previous model. Model 1.3 took 0.96 of the Akaike weight relative to Model 1.2, with a WAIC score 6.6 units lower than Model 1.1 (SD=7.69). The overlap in the standard deviation indicates that Model 1.3 overfits the data slightly, however the gap in Akaike weight shows that the addition of the 'internal evidence of failure' variable provides more benefits than hindrances in making predictions about the data. Model 1.4 below added another variable, age (G), and its interactions with the others.

(Model 1.4)

$I_i \sim \text{Ordered}(\mathbf{p})$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{USN} USN_i + \beta_{USG} USG_i + \beta_{UNG} UNG_i + \beta_{USNG} USNG_i$$

This model did not improve on Model 1.3, taking only 0.24 of the weight, and a WAIC score 2.3 units higher than Model 1.3's. However, the standard deviation of this WAIC difference was large: 6.54. This indicates that Model 1.4 overfit the data. This may have been because the parameter added – 'age' – was not informative at all, or because 'age' was informative but not enough to overcome the overfitting risk. The small amount of weight given to Model 1.4, and the large standard deviation for the difference in WAIC values, would suggest the latter. Therefore, Model 1.5 below replaced the 'internal evidence of failure' in Model 1.3 with age, to see which variable was more helpful in making predictions about the data.

(Model 1.5)

$$I_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_G G_i + \beta_{US} US_i + \beta_{UG} UG_i + \beta_{USG} USG_i$$

However, it appears that in this model too the inclusion of 'age' as a variable made the model perform worse in predicting other data from the same experiment. Compared to Model 1.3, Model 1.5 took only 0.04 of the weight. Model 1.5 had a higher WAIC value by 6.3 units, though the standard deviation of this difference was 8.76. 'Age' therefore appears to reduce the ability of the model to make predictions about the data. This would indicate that it was not a particularly important source of variation in microstructure similarity score variation under these conditions. Subsequent models for this hypothesis therefore did not include the variable of 'age'. In Model 1.6, below, I swapped out 'age' in Model 1.4 and instead introduced 'female' (F).

(Model 1.6)

$$I_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_F F_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UF} UF_i + \beta_{USN} USN_i + \beta_{USF} USF_i + \beta_{UNF} UNF_i + \beta_{USNF} USNF_i$$

This model also results in worse out-of-sample predictions than Model 1.3. It had an Akaike weight of 0.04 relative to Model 1.3, and a WAIC value higher by 6.5 (SD=5.83). I therefore also tested replacing 'failure internal' with 'female' instead:

(Model 1.7)

$$I_i \sim \text{Ordered}(\mathbf{p})$$
$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_F F_i + \beta_{US} US_i + \beta_{UF} UF_i + \beta_{USF} USF_i$$

This results in even worse predictions: an Akaike weight of 0.02 relative to Model 1.3, though with a difference of 7.7 (SD=9.4) in WAIC score compared to Model 1.3. Therefore sex differences also seem to have been relatively less helpful in predicting the effect of successful models on participants' microstructure similarity scores. The final predictor to be tested was 'attendance to the video' (T), the model for which was described as:

(Model 1.8)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_T T_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UT} UT_i + \beta_{USN} USN_i + \beta_{UST} UST_i + \beta_{UNT} UNT_i + \beta_{USNT} USNT_i$$

This model results in an Akaike weight of 0.07 relative to Model 1.3, with a higher WAIC score by 5.2 units (SD=4.49). I therefore also tried replacing 'internal failure' with 'attendance score':

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_F F_i + \beta_{US} US_i + \beta_{UF} UF_i + \beta_{USF} USF_i$$

The results for this model were even worse, with an Akaike weight of 0.01, and difference in WAIC score from Model 1.3 of 10.5 (SD=8.66). These comparisons suggest that the degree of a participant's attendance to the video was not a useful interaction variable to include in investigating the effect of a successful model on microstructure similarity scores. The model I therefore proceeded with was Model 1.3, which had the best out-of-sample deviance scores of any model thus far attempted for this hypothesis. However, Model 1.10, described below, removes 'successful' (U) as a predictor from Model 1.3, leaving just interactions between 'social' (S) and 'internal failure' (N). Comparing this model to Model 1.3 allows me to gauge the degree to which the hypothesised main predictor variable, model success, was an important influence on variation in microstructure similarity scores.

(Model 1.10)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_S S_i + \beta_N N_i + \beta_{SN} SN_i$$

Model 1.10 took 0.71 of the Akaike weight, leaving Model 1.3 on 0.29. Model 1.3 had a WAIC value 1.8 units higher than that of Model 1.10, though the standard deviation of this difference was 4.17. These scores indicated that Model 1.3 was itself somewhat overfit to the data. However, the degree of the weight which Model 1.3 took does indicate that it contained useful information, despite the overfitting. Whilst I did not exclude the 'success' variable from the model, since it was the variable that was here being examined, it was an early indicator that its effect on the outcome variable was not terribly strong.

9.2: Hypothesis 2

The first model used just the variable 'success' (U) to predict macrostructure similarity score variation.

(Model 2.1)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i$$

In the model below, the variable 'social' was introduced to distinguish successful and unsuccessful models who were either relevant or not relevant to participants' building.

(Model 2.2)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_S S_i$$

In a comparison between Models 2.1 and 2.2, Model 2.2 took 0.99 of the Akaike weight. In WAIC values, the difference between the two was 8.6 (SD=7.24). Model 2.3 below added the 'internal failure' variable (N) to this model.

(Model 2.3)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{USN} USN_i$$

This model, in comparison with Model 2.2, took 100% of the Akaike weight. The difference between the two models in WAIC was 28.1 (SD=11.37). This indicates that internal evidence of failure could have been an important mitigating factor in the effect of successful models on macrostructure similarity scores. I then added the 'age' variable (G).

(Model 2.4)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \beta_{USN} USN_i + \beta_{USG} USG_i + \beta_{UNG} UNG_i + \beta_{USNG} USNG_i$$

This model again improved on the previous one. Compared to Model 2.3, Model 2.4 took 100% of the weight, and the difference between the WAIC scores of the two was 14.1 (SD=10.66). Model 2.5 below then added 'female' to this new model.

(Model 2.5)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = {} & \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_F F_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \\ & \beta_{UF} UF_i + \beta_{USN} USN_i + \beta_{USG} USG_i + \beta_{USF} USF_i + \beta_{UNG} UNG_i + \beta_{UNF} UNF_i + \\ & \beta_{UGF} UGF_i + \beta_{USN} USN_i + \beta_{USNG} USNG_i + \beta_{USNF} USNF_i + \beta_{USGF} USGF_i + \\ & \beta_{UNGF} UNGF_i + \beta_{USNGF} USNGF_i\end{aligned}$$

This model, however, did not improve on Model 2.4 Compared to Model 2.4, Model 2.5 took none of the Akaike weight. Model 2.5 also had a higher WAIC score, by 11.7 units (SD=7.73). I therefore tried replacing 'age', as in Model 2.4, with 'female':

(Model 2.6)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = {} & \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_F F_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UF} UF_i + \beta_{USN} USN_i \\ & + \beta_{USF} USF_i + \beta_{UNF} UNF_i + \beta_{USNF} USNF_i\end{aligned}$$

Yet still this model gained no weight compared with Model 2.4. The WAIC difference between the two was 20.1, with a 12.37 standard deviation. The model I continued with was therefore Model 2.4. In Model 2.7 below, the variable 'attendance to the video' was added to Model 2.4.

(Model 2.7)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = {} & \alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_T T_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UG} UG_i + \\ & \beta_{UT} UT_i + \beta_{USN} USN_i + \beta_{USG} USG_i + \beta_{UST} UST_i + \beta_{UNG} UNG_i + \beta_{UNT} UNT_i + \\ & \beta_{UGT} UGT_i + \beta_{USN} USN_i + \beta_{USNG} USNG_i + \beta_{USNT} USNT_i + \beta_{USGT} USGT_i + \\ & \beta_{UNGT} UNGT_i + \beta_{USNGT} USNGT_i\end{aligned}$$

This model did slightly improve on Model 2.4. Model 2.7 had a higher Akaike weight, taking 0.71 to Model 2.4's 0.29. The difference in WAIC scores was just 1.8, however, with a standard deviation of 13.07. This indicates that Model 2.7 did just enough for the relative probability of useful prediction to have been tilted in its favour, overcoming the overfitting risk associated with a large number of parameters. To see if predictions could be improved further, I attempted to replace 'age' in Model 2.4 with the 'attendance to the video variable:

(Model 2.8)

A_i~Ordered(**p**)

$$\text{logit}(p_k)=\alpha_k + \beta_U U_i + \beta_S S_i + \beta_N N_i + \beta_T T_i + \beta_{US} US_i + \beta_{UN} UN_i + \beta_{UT} UT_i + \beta_{USN} USN_i + \beta_{UST} UST_i + \beta_{UNT} UNT_i + \beta_{USNT} USNT_i$$

However, this model gained no Akaike weight in comparison with either Model 2.4 or Model 2.7. 'Age' seems to have been a crucial predictor. The model with the highest WAIC score considered so far for this hypothesis was Model 2.7. It was therefore with this model that I continue. Model 2.9, below, describes interactions of variables for macrostructure similarity scores in this condition, but without model success being included as a predictor variable.

(Model 2.9)

A_i~Ordered(**p**)

$$\text{logit}(p_k)=\alpha_k + \beta_S S_i + \beta_N N_i + \beta_G G_i + \beta_T T_i + \beta_{SN} SN_i + \beta_{SG} SG_i + \beta_{ST} ST_i + \beta_{NG} NG_i + \beta_{NT} NT_i + \beta_{GT} GT_i + \beta_{SN} SN_i + \beta_{SNG} SNG_i + \beta_{SNT} SNT_i + \beta_{SGT} SGT_i + \beta_{NGT} NGT_i + \beta_{SNGT} SNGT_i$$

This model attains a higher Akaike weight than Model 2.7 at 0.62. The difference between the two models' WAIC scores was 1 (SD=14.63). This indicates that model success may not have been a very good predictor of participants' macrostructure similarity scores in this condition, but the large standard deviation made this assessment uncertain.

9.3: Hypothesis 3

Model 3.1, below, used internal evidence of failure (N) alone to predict variation in microstructure similarity scores.

(Model 3.1)

I_i~Ordered(**p**)

$$\text{logit}(p_k)=\alpha_k + \beta_N N_i$$

The second predictor variable I added to this model was model sociality (S), below.

(Model 3.2)

I_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_N N_i + \beta_S S_i + \beta_{NS} NS_i$

This marks an improvement on Model 3.1, with Model 3.2 taking 100% of the Akaike weight and a lower WAIC score by 17.8 units (SD=9.42). I thus added another variable to this model: model success (U).

(Model 3.3)

I_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_N N_i + \beta_S S_i + \beta_U U_i + \beta_{NS} NS_i + \beta_{NU} NU_i + \beta_{NSU} NSU_i$

This new model did not perform as well as Model 3.2. Model 3.3 gained only 0.04 of the weight, and had a higher WAIC value by 6.2 units (SD=1.87). This indicates that model success did not add useful information in predicting microstructure similarity scores in this context. To further test this, I replaced the 'social' variable in Model 3.2 with 'successful':

(Model 3.4)

I_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_N N_i + \beta_U U_i + \beta_{NU} NU_i$

α_k~Normal(0,10)

β_N~Normal(0,10)

β_U~Normal(0,10)

β_{NU}~Normal(0,10)

In comparison with Model 3.2, Model 3.4 took none of the Akaike weight. It had a WAIC score 21.6 units higher than Model 3.2's, with a standard deviation of 9.51. Since the 'successful' predictor appears to reduce the inferential power of the model both when included in interaction with the sociality of the model and when replacing the sociality of the model, I did not continue adding it as a predictor to other models for this hypothesis. Model 3.5, below, thus added a variable for the age of the participants to Model 3.2.

(Model 3.5)

I_i~Ordered($\mathbf{p}$)

logit(p_k)=$\alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NSG} NSG_i$

This model also did not improve on Model 3.2. In a comparison between the two, Model 3.5 gained just 0.34 of the weight, to Model 3.2's 0.66. However, the difference in WAIC scores was small, at 1.3 and with a standard deviation of

4.96. This indicates uncertainty about which model provides the best predictions of future data. I therefore tried replacing 'social' in Model 3.2 with 'age':

(Model 3.6)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_G G_i + \beta_{NG} NG_i$$

Model 3.6 gained no weight relative to Model 3.2, and the difference in WAIC scores was 20.8 (SD=10.2). This indicates that the 'age' predictor only had value for explaining variation in the data once the 'social' predictor was taken into account. However, it seems the value that the 'age' predictor added was not enough to overcome the overfitting which the addition of another variable brings, leading the greater probability of useful prediction to lie with Model 3.2. Despite the uncertainty in the WAIC comparison between Models 3.2 and 3.5, I therefore continued with Model 3.2: the simpler model which nevertheless gained the greater probability of better predicting new results from the same experimental process. Model 3.7 below thus added 'female' to Model 3.2:

(Model 3.7)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_F F_i + \beta_{NS} NS_i + \beta_{NF} NF_i + \beta_{NSF} NSF_i$$

Compared to Model 3.2, Model 3.7 gained only 0.22 of the Akaike weight, compared to Model 3.2's 0.78. However, the difference between their WAIC scores (2.5) was not greater than the standard deviation of the difference (4.3), indicating uncertainty in the relative usefulness of the models. I therefore replaced 'social' in Model 3.2 with 'female'.

(Model 3.8)

$$I_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_F F_i + \beta_{NF} NF_i$$

A comparison between these two models reveals that Model 3.8 took none of the weight. The difference between their WAIC scores was also large, at 19.9 (SD=10.44). This indicates that, in interaction with 'internal evidence of failure' only, 'female' was far less useful a predictor of microstructure similarity score variation than was 'social'. Further, the comparison with Model 3.7 indicates that there was likely a lower probability that a model with interactions between 'internal evidence of failure', 'social', and 'female' made better predictions than a simpler model with interactions just between 'internal evidence of failure' and 'social'. The model I therefore continued with was still Model 3.2. Model 3.9 below added interactions with 'attendance to the video' (T) to Model 3.2.

(Model 3.9)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_T T_i + \beta_{NS} NS_i + \beta_{NT} NT_i + \beta_{NST} NST_i$

However, this model produced a similar result to Models 3.5 and 3.7. In comparison with Model 3.2, Model 3.9 took 0.28 of the weight. The difference in WAIC scores was 1.9 (SD=3.81). Thus again there was uncertainty in which model did produce the better predictions, though the greater probability lies with Model 3.2. Thus Model 3.10 swapped 'social' in Model 3.2 for 'attendance to the video'.

(Model 3.10)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_T T_i + \beta_{NT} NT_i$

The comparison between these two reveals that Model 3.10 took none of the Akaike weight. The difference between the two models' WAIC scores was 16.6, with a standard deviation of 10.44. This indicates that a participant's attendance to the experimental video was more clearly not useful as a predictor when it was not considered in interaction with the participant's age. I therefore continued with the model that combines the simplest model description with the biggest probability of the lowest WAIC and highest weight values: Model 3.2. The model below removes the 'attendance to the video' variable from Model 3.2.

(Model 3.11)

$I_i \sim \text{Ordered}(\mathbf{p})$

$\text{logit}(p_k) = \alpha_k + \beta_S S_i$

Model 3.11 took only 0.02 of the weight, compared to Model 3.2's 0.98. However, there was a slight overlap of the standard deviation in the difference in WAIC scores: 7.8 (SD=7.83). This means that most of the probability was for Model 3.2 making better predictions than Model 3.11. This therefore suggests that the 'internal evidence of failure' variables did provide some useful information for learning about variation in microstructure similarity scores. I therefore still continued with Model 3.2.

9.4: Hypothesis 4

Model 4.1, below, used internal evidence of failure only to predict variation in macrostructure similarity scores.

(Model 4.1)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i$$

I next added the 'social' variable to this model:

(Model 4.2)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_{NS} NS_i$$

Model 4.2 took 100% of the Akaike weight, the difference between their WAIC scores being 17.9 (SD=9.03). This indicates that the sociality of the model was a useful piece of data for understanding the effect of internal evidence of failure on macrostructure similarity scores. Model 4.3 below added model success to this.

(Model 4.3)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_U U_i + \beta_{NS} NS_i + \beta_{NU} NU_i + \beta_{NSU} NSU_i$$

Compared to Model 4.2, Model 4.3 took 0.35 of the weight. The difference in WAIC scores was 1.2, with a large 4.84 standard deviation. This indicates uncertainty in which model was more useful, though there was a greater probability that it was Model 4.2 than Model 4.3. The model below thus swapped out 'social' in Model 4.2 for 'successful'.

(Model 4.4)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_{NS} NS_i$$

Compared with Model 4.2, Model 4.4 receives no Akaike weight. The difference between their WAIC scores was 16.8 (SD=9.83). Interactions with model success, alone, therefore seems relatively unimportant for understanding variation in macrostructure similarity scores here. I therefore continued with Model 4.2, to which I added a variable for participant age.

(Model 4.5)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NSG} NSG_i$$

Model 4.5, in comparison with Model 4.2, took all of the Akaike weight, the difference between their WAIC values being 12.8 (SD=9.43). This indicates that the participants' ages were a useful source of information for understanding the influence of internal evidence of failure on macrostructure similarity ratings.

I therefore continued with Model 4.5 by adding the 'female' variable to it.

(Model 4.6)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_F F_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NF} NF_i + \beta_{NSG} NSG_i + \beta_{NSF} NSF_i + \beta_{NGF} NGF_i + \beta_{NSGF} NSGF_i$$

The addition of interactions with 'age' appears to reduce the effectiveness of the model. In comparison with Model 4.5, Model 4.6 took just 0.04 of the weight. The difference between the WAIC values of the two models was 6.6 (SD=5.62). Model 4.7, below, instead tries replacing 'age' in Model 4.5 with 'female'.

(Model 4.7)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_F F_i + \beta_{NS} NS_i + \beta_{NF} NF_i + \beta_{NSF} NSF_i$$

In a comparison with Model 4.5, Model 4.7 took none of the Akaike weight. The difference in WAIC values between the two was 14.9, with a standard deviation of 10.67. This indicates that the effect of adding 'female' reduces the ability of the model to predict new data even when other parameters were removed to guard against overfitting. I therefore continued with Model 4.5. Model 4.8 added 'attendance to the video' to Model 4.5.

(Model 4.8)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_T T_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NT} NT_i + \beta_{NSG} NSG_i + \beta_{NST} NST_i + \beta_{NGT} NGT_i + \beta_{NSGT} NSGT_i$$

The addition of this variable appears to increase the model's ability to make predictions about future similar data. Compared with Model 4.5, Model 4.8 took

0.98 of the weight, the difference between their WAIC scores being 7.7 though with a standard deviation of 10.83. This suggests some uncertainty, though the greater probability for better predictions lies with Model 4.8. Compared with another model, Model 4.9 (not described here), in which 'age' was replaced by 'attendance to the video', Model 4.8 took all of the weight, with a difference in WAIC scores of 27.8 (SD=14.6). Out of the models reviewed so far, Model 4.8 thus remains the most likely to make useful predictions about new data from a similar experimental process. However, the role of model success was still unclear, since the WAIC difference between models 4.2 and 4.3 was overshadowed by the standard deviation of the difference. I therefore added 'success' back into Model 4.8 to see whether it results in better or worse predictions.

(Model 4.10)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\begin{aligned}\text{logit}(p_k) = {} & \alpha_k + \beta_N N_i + \beta_S S_i + \beta_G G_i + \beta_T T_i + \beta_U U_i + \beta_{NS} NS_i + \beta_{NG} NG_i + \beta_{NT} NT_i + \\ & \beta_{NU} NU_i + \beta_{NSG} NSG_i + \beta_{NST} NST_i + \beta_{NSU} NSU_i + \beta_{NGT} NGT_i + \beta_{NGU} NGU_i + \\ & \beta_{NTU} NTU_i + \beta_{NSGT} NSGT_i + \beta_{NSGU} NSGU_i + \beta_{NGTU} NGTU_i + \beta_{NSGTU} NSGTU_i\end{aligned}$$

Here the effect of adding 'successful' to the model was still not clearly negative. Compared to Model 4.8, Model 4.10 took 0.16 of the Akaike weight. But the difference in WAIC scores between the two was 3.3, with a standard deviation of 9.23. The model that I continued analysis with was Model 4.8. This was because the comparisons indicated that there was a greater probability of better predictions with this model than Model 4.10, and because Model 4.8 was simpler and can explain the variation in the data with as much reliability, if nor more, as Model 4.10 despite having fewer parameters. Model 4.11, below, removes 'internal evidence of failure' from Model 4.8.

(Model 4.11)

$$A_i \sim \text{Ordered}(\mathbf{p})$$

$$\text{logit}(p_k) = \alpha_k + \beta_S S_i + \beta_G G_i + \beta_T T_i + \beta_{SG} SG_i + \beta_{ST} ST_i + \beta_{GT} GT_i + \beta_{SGT} SGT_i$$

Compared to Model 4.8, Model 4.11 gained none of the Akaike weight. The difference between the WAIC scores of the two models was 17.4, with a standard deviation of 12.63. This indicates that the main predictor variable, the degree of internal evidence of failure, did have an interesting influence on macrostructure similarity scores.

Appendix 10: Further analyses of data for Chapter 6

10.1: Hypothesis 1

Changing an asocial model (Figure 34's graphs A and B, page 98) into a social model (Figure 34's graphs C and D) inverted the effect of the successful model on microstructure similarity scores both for the weakly negative effect of the successful model in the asocial condition with low internal evidence of failure (graph A), and the weakly positive effect of the successful model with high internal evidence of failure (graph B). Both of these negative and positive relationships were reversed when the model was social rather than asocial.

Model 11 predicted, in Figure 34, that participants with the asocial model and low internal evidence of failure (graph A) would show a slightly negative relationship between the success of the model and microstructure similarity scores. However, Figure 34 also predicted that participants with the asocial model and high internal evidence of failure (graph B) would show a slightly positive relationship between model success and microstructure similarity scores. The same inversion effect was present with the social model, except there it was the participants with low internal evidence of failure (graph C) that showed the slightly positive relationship between model success and microstructure similarity scores, and participants with high internal evidence of failure (graph D) that showed the slightly negative one.

10.2: Hypothesis 2

The influence of high internal evidence of failure, rather than low internal evidence of failure, on the effect of changing an unsuccessful model into a successful model appeared variable. In Figure 36 (page 102), among younger children exhibiting low attendance to the video, higher internal evidence of failure caused a lack of effect of social model success on macrostructure similarity with lower internal evidence of failure to become a negative effect of model success on macrostructure similarity (between graphs C and D). In Figure 37 (page 103), among older children exhibiting low attendance to the video, higher internal evidence of failure seemed merely to conserve the positive effect of a successful model on macrostructure similarity from participants with low internal evidence of failure (between graphs C and D). In Figure 38 (page 103), with younger children exhibiting high attendance to the video, higher internal evidence of failure caused a lack of effect of model success on macrostructure similarity in graph C to become a positive effect of model success on macrostructure similarity in graph D. In Figure 39 (page 104), with older children exhibiting higher attendance to the video, higher internal evidence of failure (graph D) appeared to reverse the positive effect of a successful model with low internal evidence of failure (in graph C). In the asocial model conditions, the effect of higher internal evidence of failure was often not the same as in the social conditions. This was more so with children exhibiting higher attendance to the video. In Figure 38, with younger children exhibiting higher attendance to the video, higher internal evidence of failure appeared to reverse the direction of the effect of a successful asocial model on macrostructure similarity (between graphs C and D). This reversal of the effect of a successful asocial model was also visible in Figure 39's graphs C and D, with older children exhibiting higher attendance to the video. In Figures 36 and 37, where younger and older children both exhibited lower attendance to the video, then the impact of higher internal evidence of failure, graph D, was merely to maintain the direction of the relationship between model success and macrostructure similarity scores found in the 'C' graphs.

Comparing the conditions with low participant age, Figures 36 and 38 had very similar effects of social and asocial model success when internal evidence of failure was low (graphs A and C in each Figure). When internal evidence of failure was high in both Figure 36 and Figure 38 (graphs B and D), the effects of social model success were the reverse of when internal evidence of failure was low, though only Figure 38 (with higher attendance to the video) showed high internal evidence of failure to reverse the effect of asocial model success as well. A similar pattern can be found also in Figures 37 and 39, with older participant age. The difference in the effects of the successful model between social and asocial conditions appeared very similar between older children exhibiting greater and lesser attendance to the video, when internal evidence of failure was low (i.e., in graphs A and C). The older children did show similar effects of model success in the social and asocial conditions when internal evidence of failure was high (i.e., there was similarity in the 'B' graphs of Figures 37 and 39, as well as similarity in the 'D' graphs of Figures 37 and 39). However, when internal evidence of failure was low, the direction of the effect of the successful model on macrostructure similarity was different between Figures 37 and 39. In Figure 37, with low attendance to the video (graphs A and C), the effect of the successful model on macrostructure similarity remained positive with high internal evidence of failure (graphs B and D), whilst in Figure 39, with high attendance to the video, the effect of the successful model on macrostructure similarity became negative with high internal evidence of failure. The influence of age on the effect of model success on macrostructure similarity thus appeared to have been highly dependent on the influence of the other variables.

The influence of high attendance to the video, contrarily, appeared quite uniform in the asocial conditions (graphs A and B). Across Figures 36 and 37, with lower attendance to the video and lower and higher participant ages, the impact of the asocial model on model success' effect on macrostructure similarity was consistent across both lower and higher internal evidence of failure (graphs A and B). This was also true between Figures 38 and 39, with high participant attendance to the video and lower and higher participant ages. Conversely, between Figures 36 and 38, and Figures 37 and 39, the direction of the effect of the asocial successful model where participant evidence of failure was high was reversed (i.e., between graphs A and B). The impact of higher attendance to the video on the effect of the successful model was again more variable in the social model conditions (graphs C and D). The effect of the successful social model was similar for participants exhibiting low internal evidence of failure (graph C) across Figures 36 and 37, with lower attendance to the video and lower and higher participant age. However the effect of the successful model was also reversed in direction for children with higher internal evidence of failure (graph D) in the same Figures. The same story held true in comparing the effect of the successful social model in Figures 38 and 39, with higher attendance to the video and lower and higher participant age (graphs C and D respectively).

10.4: Hypothesis 4

The influence of the social, rather than asocial, model on high internal evidence of failure's effect on macrostructure similarity scores appears to have been to reduce its negative effect when the model was asocial. In the asocial conditions (graphs A and B of Figures 43 and 44, pages 116 and 117 respectively), there was a reliably negative effect of high internal evidence of failure on macrostructure similarity. While this negative relationship was present in two social model conditions (in Figure 43's graph C, with younger children exhibiting lower attendance to the video, and Figure 44's graph D, with older children exhibiting higher attendance to the video), in the two other social conditions this negative effect was removed by the apparent influence of, on the one hand, high child age with low video attendance (graph D in Figure 43), and on the other hand, low child age with high video attendance (graph C in Figure 44). The social model thus enables these interactions to reduce the negative relationship between higher internal evidence of failure and macrostructure similarity with both: (1) an asocial model with high child age and low video attendance, and (2) an asocial model with low child age and high video attendance.

The influence of higher participant age on the effect of higher internal evidence of failure appears to have been complex. In the context of the asocial model, across low and high attendance to the video (Figures 43 and 44), higher participant age (graphs B and D) appears to cause greater macrostructure similarity with both low or high internal evidence of failure (comparing the left and right hand sides of each graph), though it appears not to change the negative relationship between increased internal evidence of failure and macrostructure similarity scores. In the context of a social model (graphs C and D), increased age appears to have two opposite influences on the effect of internal evidence of failure on macrostructure similarity. When demonstrating low attendance to the video (Figure 43), increased age in graph D weakens the negative effect of high internal evidence of failure on macrostructure similarity (in graph C) to such a degree that high internal evidence of failure appears to have no directional influence on macrostructure similarity scores. However, when demonstrating high attendance to the video (Figure 44), increased age appears to change high internal evidence of failure from having no effect on macrostructure similarity in graph C to having a clearly negative effect on macrostructure similarity scores in graph D.

Higher participant attendance to the experimental video appears to have had little impact on the effect of higher internal evidence of failure on macrostructure similarity scores with an asocial model (graphs A and B). Across Figures 43 and 44, with the asocial model, the negative effect of internal evidence of failure on macrostructure similarity was consistent. Yet like for the impact of increased participant age above, the role of increased attendance to the video was variable in with the social model (graphs C and D). In the social condition with low participant age (graph C), the change from low to high attendance to the video causes the relationship between internal evidence of failure and macrostructure similarity to change from negative to neutral. In the social condition with high participant age (graph D), the change from low to high attendance to the video causes the relationship between internal evidence of failure and macrostructure similarity to change from neutral to negative. Thus it appears that in these social model conditions, the relationship between internal evidence of failure and macrostructure similarity was complex and dependent on a combination of other variables.

Appendix 11: Lay summary of the thesis

The purpose of my research was to study how children copied from others in play. A key part of play is that children's activities are open-ended and not determined by other people. I ran an experiment in which 565 primary school aged children built with wooden blocks. Some of these children had the possibility to copy from someone else who was also building. I then tried to manipulate different conditions to cause children to build things which were more or less similar to what this other person built. I did this in order to find out two things. The first was whether children copied differently when the task was open-ended (where children were told to build whatever they thought was best) compared to when the task was close-ended (where children were told to build the tallest tower). I found that there were some differences. For example, the children who were told to build a tall tower copied the other person less when this other person was worse at building compared to when this other person was better at building. However, children who were told to build whatever they liked did not copy the other person less when the other person's building was unsuccessful. My second aim was to see whether children's copying was more flexible in the open-ended task than it was in the close-ended task. The data, however, did not support this idea. In addition to these results, I found that children who were older than seven copied the other person more than children who were younger than seven, and that girls tended to copy the other person slightly more than boys. Overall, my experiment shows that the way in which children's activities are framed, as either more close- or open-ended, can have effects on how children react to information provided by others. This research therefore indicates that play may be a special context for children, in which they react to social information differently compared to how they copy when they are given a specific goal to achieve.

Works Cited

Amabile, T. M. *Creativity in Context.* Boulder: Westview Press, 1996.

Astington, J. W. "Intention in the child's theory of mind." In *Children's theories of mind: Mental states and social understanding*, edited by D. Frye and C. Moore, 157-172. Hillsdale: Erlbaum, 1991.

Astington, J. W., and A. Gopnik. "Theoretical explanations of children's understanding of the mind." *British Journal of Developmental Psychology* 9 (1991): 7-31.

Athey, I. "Contributions of Play to Development." In *Child's Play: Developmental and Applied*, edited by T. D. Yawkey and A. D. Pellegrini, 9-28. Hillsdale: Erlbaum, 1984.

Baer, D. M., R. F. Peterson, and J. A. Sherman. "The development of imitation by reinforcing behavioural similarity to a model." *Journal of the Experimental Analysis of Behavior* 10, no. 5 (September 1967): 405-416.

Bandura, A. *The social foundations of thought and action: A social cognitive theory.* Englewood Cliffs: Prentice-Hall, 1986.

Barrett, H. C., C. D. Peterson, and W. E. Frankenhuis. "Mapping the Cultural Learnability Landscape of Danger." *Child Development* 87, no. 3 (May/June 2016): 770-781.

Barrett, L., S. P. Henzi, and D. Lusseau. "Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals." *Philosophical Transactions of the Royal Society B* 367 (2012): 2108-2118.

Bates, E., V. Carlson-Lunden, and I. Bretherton. "Perceptual aspects of tool using in infancy." *Infant Behavior and Development* 3 (January 1980): 127-140.

Bateson, P, and P. Martin. *Play, Playfulness, Creativity and Innovation.* Cambridge: Cambdrige University Press, 2013.

Bateson, P. "Play, Playfulness, Creativity and Innovation." *Animal Behavior and Cognition* 1, no. 2 (2014): 99-112.

Beauchamp, G., and A. Kacelnik. "Effects of the knowledge of partners on learning rates in zebra finches." *Animal Behavior* 41, no. 2 (February 1991): 247-253.

Beck, S. R. et al. "Making tools isn't child's play." *Cognition* 119 (2011): 301-306.

Bekkering, H., and W. Prinz. "Goal representations in imitative actions." In *Imitation in animals and artifacts*, edited by K. Dautenhahn and C. Nehaniv, 553-572. Cambridge: MIT Press, 2002.

Bellagamba, F., and M. Tomasello. "Re-enacting intended acts: comparing 12- and 18-month-olds." *Infant Behaviour and Development*, 1999: 277-282.

Berger, J. "Arousal increases social transmission of information." *Psychological Science* 22, no. 7 (2011): 891-893.

Besemer, S. P., and K. O'Quin. "Confirming the three-factor creative product analysis matrix model in an American sample." *Creativity Research Journal* 12 (1999): 287-296.

Bijovet-van den Berg, C. J. M. "Children's Ability to Generate Novel Actions." PhD Thesis: University of Stirling, October 2013.

Birch, S., S. Vautheir, and P. Bloom. "Three- and four-year-olds spontaneously use others' past performance to guide their learning." *Cognition* 107 (2008): 1018-1034.

Biro, D., M. Haslam, and C. Rutz. "Tool use as adaptation." *Philosophical Transactions of the Royal Society B* 368 (2013): 20120408.

Bonawitz, E. et al. "The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery." *Cognition* 120 (2011): 322-330.

Botting, J., E. van de Waal, and L. Rendell. "Comparing and Contrasting Primate and Cetacean Culture." In *The Handbook of Culture and Biology*, 105-128. Oxford: John Wiley & Sons, 2018.

Burdett, E. R. R. et al. "The interaction of social and perceivable-causal factors in shaping 'over-imitation'." *Cognitive Development* 47 (2018): 8-18.

Boyd, R., and P. J. Richerson. *Culture and the Evolutionary Process.* Chicago: The University of Chicago Press, 1985.

Boyd, R., and P. J. Richerson. "Why Does Culture Increase Human Adaptability?" *Ethology and Sociobiology* 16 (1995): 125-143.

Boyette, A. H., and B. S. Hewlett. "Teaching in Hunter-Gatherers." *Review of Philosophy and Psychology* 9, no. 4 (2018): 771-797.

Brahms, L., and K. Crowley. *Learning to Make in the Museum: The Role of Maker Educators.* Vol. 1, in *Makeology in K-12, Higher, and Informal Education: The Maker Movement and the Future of Learning*, edited by K. A. Pepler, E. R. Halverson and Y. B. Kafai, 15-29. New York: Routledge, 2016.

Brand, C. O., G. R. Brown, and C. P. Cross. "Sex differences in the use of social information emerge under conditions of risk." *PeerJ* 6 (2018): e4190.

Brown, J. S., A. Collins, and P. Daguid. "Situated Cognition and the Culture of Learning." *Educational Researcher* 18, no. 1 (January-February 1989): 32-42.

Bullock, M., R. Gelman, and R. Baillargeon. "The development of causal reasoning." In *The developmental psychology of time*, edited by W. J. Friedman, 209-254. New York: Academic Press, 1982.

Burkner, P.-C., and M. Vuorre. "Ordinal Regression Models in Psychological Research: A Tutorial." (PsyArXiv Preprints) February 2018.

Burnham, K. P., and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd. New York: Springer-Verlag, 2002.

Caldwell, C. A., and A. E. Millen. "Conservatism in laboratory microsocieties: unpredictable payoffs accentuate group-specific traditions." *Evolution and Human Behavior* 31 (2010): 123-130.

Caldwell, C. A., and A. E. Millen. "Social learning mechanisms and cumulative cultural evolution: Is imitation necessary?" *Psychological Science* 20, no. 12 (December 2009): 1478-1483.

Call, J., and M. Carpenter. "Three Sources of Information in Social Learning." In *Imitation in Animals and Artifacts*, edited by K. Dautenhahn and C. L. Nehaniv, 211-228. London: MIT Press, 2002.

Call, J., and M. Tomasello. "Use of social information in the problem solving of orangutans (Pongo pygmaeus) and in human children (Homo sapiens)." *Journal of Comparative Psychology* 109 (1995): 308-320.

Carpenter, B. et al. "Stan: A Probabilistic Programming Device." *Journal of Statistical Software* 76, no. 1 (January 2017).

Carpenter, M., J. Call, and M. Tomasello. "Twelve- and 18-month-olds copy actions in terms of goals ." *Developmental Science* 8, no. 1 (2005): F13-F20.

Carpenter, M., N. Akhtar, and M. Tomasello. "Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions." *Infant Behavior & Development* 21, no. 2 (1998): 315-330.

Carr, K. "Childhood Innovation: Development, Facilitators, and Individual Variation." *PhD Thesis.* Durham University, 2016.

Carr, K., R. Kendal, and E. Flynn. "Imitate or innovate? Children's innovation is influenced by the efficacy of observed behaviour." *Cognition* 142 (September 2015): 322-332.

Carr, K., R. L. Kendal, and E. G. Flynn. "Eureka!: What Is Innovation, How Does It Develop, and Who Does It?" *Child Development* 87, no. 5 (2016): 1505-1519.

Chappell, A. et al. "The development of tool manufacture in humans: what helps young children make innovative tools?" *Philosophical Transactions of the Royal Society B* 368 (2013): 20120409.

Christiaans, H. H. C. M. "Creativity as a design criterion." *Creativity Research Journal* 14 (2002): 41-54.
Clay, Z., and C. Tennie. "Is Overimitation a Uniquely Human Phenomenon? Insights from Human Children as Compared to Bonobos." *Child Development*, 2017: Avance online publication.

Clay, Z., H. Over, and C. Tennie. "What drives young children to over-imitate? Investigating the effects of age, context, action type, and transitivity." *Journal of Experimental Child Psychology* 166 (2018): 520-534.

Clement, F., M. Koenig, and P. Harris. "The Ontogenesis of Trust." *Mind & Language* 19, no. 4 (2004): 360-379.

Cohen, J. "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement* 20, no. 1 (April 1960): 37-46.

Cohen, J. "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." *Psychological Bulletin* 70, no. 4 (1968): 213-220.

Cook, C., N. D. Goodman, and L. E. Schultz. "Where science starts: Spontaneous experiments in children's exploratory play." *Cognition* 120 (2011): 341-349.

Csibra, G., and G. Gergely. "Natural pedagogy as evolutionary adaptation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, no. 1567 (2011): 1149-1157.

Cutting, N., I. A. Apperly, and S. R. Beck. "Why do children lack the flexibility to innovate tools?" *Journal of Experimental Child Psychology* 109 (2011): 497-511.

Darby, C. L., and A. J. Riopelle. "Observational learning in the rhesus monkey." *Journal of Comparative and Physiological Psychology* 52, no. 1 (1959): 94-98.

Davenport, T., and J. Beck. *The attention economy: Understanding the new economy of business.* Cambridge: Harvard Business School Press, 2001.

de Waal, F. B. M. *Are we smart enough to understand how smart animals are?* London: Granta, 2016.

Dean, L. G. et al. "Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture." *Science* 335 (March 2012): 1114-1118.

Defeyter, M. A., S. E. Avons, and T. C. German. "Developmental changes in information central to artifact representation: evidence from 'functional fluency' tasks." *Developmental Science* 10 (2007): 538-546.

DiYanny, C., D. Nini, and W. Rheel. "Looking good versus doing good: Which factors take precedence when children learn about new tools." *Experimental Child Psychology* 110 (2011): 575-591.

Dove, G. "How to go beyond the body: an introduction." *Frontiers in Psychology* 6 (May 2015): Article 660.
Ehrlinger, J., and D. Dunning. "How chronic self-views influence (and potentially mislead) estimates of performance." *Journal of Personality and Social Psychology* 84 (2003): 5-17.

Enquist, M., K. Eriksson, and S. Ghirlanda. "Critical Social Learning: A Solution to Rogers's Paradox of Nonadaptive Culture." *American Anthropologist* 109, no. 4 (2007): 727-734.

Ervin, S. J. K. et al. "Estrogen involvement in social behavior in rodents: rapid and long-term actions." *Hormones and Behavior* 74 (August 2015): 53-76.

Evans, C. L. et al. "Selective copying of the majority suggests children are broadly "optimal-" rather than "over-" imitators." *Developmental Science* Early View (2017): e12637.

Fausto, C., and C. Valentina. "Types of abduction in tool behavior." *Phenomenology and the Cognitive Sciences* 16, no. 2 (April 2017): 255-273.

Feldman, M. W., K. Aoki, and J. Kumm. "Individual versus social learning: Evolutionary analysis in a fluctuating environment." *Anthropological Science* 104 (1996): 209-232.

Festinger, L. "A Theory of Social Comparison Processes." *Human Relations* 7 (1954): 114-140.

Fitch, W. T. "On externalization and cognitive continuity in language evolution." *Mind & Language* 32 (2017): 597-606.

Flynn, E. G., K. N. Laland, R. L. Kendal, and J. R. Kendal. "Developmental niche construction." *Developmental Science* 16, no. 2 (2013): 296-313.

Flynn, E. "Investigating children as cultural magnets: Do young children transmit redundant information along diffusion chains?" *Philosophical Transactions of the Royal Society B* 363 (2008): 3541-3551.

Flynn, E., and A. Whiten. "Imitation of hierarchical structure versus component details of complex actions by 3- and 5-year-olds." *Journal of Experimental Child Psychology* 101 (2008): 228-240.

Flynn, E., T. Turner, and L-A. Giraldeau. "Selectivity in social and asocial learning: investigating the prevalence, effect and development of children's social learning." *Philosophical Transactions of the Royal Society B* 371 (2016): 20150189.

Fogarty, L. "Cultural complexity and evolution in fluctuating environments." *Philosophical Transactions of the Royal Society B* 373, no. 1743 (April 2018): 20170063.

Freeman, N. K. "Preschoolers' Perceptions of Gender Appropriate Toys and their Parents' Beliefs About Genderized Behaviors: Miscommunication, Mixed Messages, or Hidden Truths?" *Early Childhood Education Journal* 34, no. 5 (April 2007): 357-366.

Frost, J. L., S. C. Wortham, and S. C. Reifel. *Play and Child Development.* 4th. London: Pearson, 2012.

Garvey, C. *Play.* 2nd. Cambridge: Harvard University Press, 1990.

German, T. C., and M. A. Defeyter. "Immunity to functional fixedness in young children." *Psychonomic Bulletin & Review* 7 (2000): 707-712.

Gibson, J. J. *The ecological approach to visual perception.* London: Erlbaum, 1986.

Giraldeau, L.-A., T. J. Valone, and J. J. Templeton. "Potential disadvantages of using socially-acquired information." *Philosophical Transactions of the Royal Society B* 357 (2002): 1559-1566.

Golinkoff, R. M., K. Hirsh-Pasek, and D. G. Singer. "Why Play = Learning: A Challenge for Parents and Educators." In *Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth*, edited by D. G. Singer, R. M. Golinkoff and K. Hirsh-Pasek, 3-12. Oxford: Oxford University Press, 2006.

Gollwitzer, P. M. *Mindset Theory of Action Phases.* Vol. 1, in *Handbook of Theories of Social Psychology*, edited by P. A. M. van Lange, A. W. Kruglanski and E. T. Higgins, 526-545. Los Angeles: Sage, 2012.

Goodwin, C., and M. H. Goodwin. "Participation." In *A Companion to Linguistic Anthropology*, edited by A. Duranti, 222-244. Oxford: Blackwell, 2004.

Gopnik, A. "Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications." *Science* 337 (September 2012): 1623-1627.

Gray, P. *Free to Learn: Why Unleashing the Instinct to Play will make our Children Happier, More Self-Relient, and Better Students for Life.* New York: Basic Books, 2013.

Hauser, M. D. et al. "The Mystery of Language Evolution." *Frontiers in Psychology* 5 (May 2014): 401.

HEFCE, SFC, HEFCW, and DELNI. "Assessment Framework and Guidance on Submissions (02.2011 Updated to Include Addendum Published in January 2012)." In *REF2014*. London, 2012.

Henrich, J. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter.* Princeton: Princeton University Press, 2016.

Henrich, J., S. J. Heine, and A. Norenzayan. "The weirdest people in the world?" *Behavioral and Brain Sciences* 33 (2010): 61-135.

Herbert, M. J., and C. M. Harsh. "Observational learning by cats." *Journal of Comparative Psychology* 37, no. 2 (1944): 81-95.

Heyes, C. M. "Born pupils? Natural pedagogy and cultural pedagogy." *Perspectives on Psychological Science* 11, no. 2 (March 2016): 280-295.

Heyes, C. M. "Causes and consequences of imitation." *Trends in Cognitive Sciences* 5, no. 6 (June 2001): 253-261.

Heyes, C. M. "Social learning in animals: categories and mechanisms." *Biological Reviews* 69, no. 2 (May 1994): 207-231.

Heyes, C. M. "What's Social About Social Learning?" *Journal of Comparative Psychology* 126, no. 2 (2012): 193-202.

Heyes, C. M. "When does social learning become cultural learning?" *Developmental Science* 20, no. 2 (March 2017): e12350.

Hood, B. M. "Gravity rules for 2- to 4-year-olds?" *Cognitive Development* 10 (1995): 577-598.

Hopper, L. M. et al. "Effect of personality, age, sex, and estrous state on chimpanzee problem-solving success." *Animal Cognition* 17, no. 4 (2014): 835-847.

Horner, V., and A. Whiten. "Learning from others' mistakes? Limits on understanding a trap-tube task by young chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*)." *Journal of Comparative Psychology* 121, no. 1 (February 2007): 12-21.

Huang, C. T., and T. Charman. "Graduations of emulation learning in informants' imitation of actions on objects." *Journal of Experimental Child Psychology* 92 (2005): 276-302.

John, E. R. et al. "Observation learning in cats." *Science* 159 (1968): 1489-1491.

Kabadayi, C., and M. Osvarth. "Ravens parallel great apes in flexible planning for tool-use and bartering." *Science* 357, no. 6347 (July 2017): 202-204.

Kalish, M. L., T. L. Griffiths, and S. Lewandowsky. "Iterated learning: intergenerational knowledge learning reveals inductive biases." *Psychonomic Bulletin & Review* 14, no. 2 (April 2007): 288-294.

Kendal, R. L. et al. "Designing for creativity and innovation in informal science learning." *Informal Learning Review* 137 (2016): 20-24.

Kendal, R. L. et al. "Social Learning Strategies: Bridge-Building between Fields." *Trends in Cognitive Sciences* 22, no. 7 (July 2018): 651-665.

Kendal, R. L. et al. "Trade-Offs in the Adaptive Use of Social and Asocial Learning." *Advances in the Study of Behavior* 35 (2005): 333-379.

Krupenye, C. et al. "Great apes anticipate that other individuals will act according to false beliefs." *Science* 354, no. 6308 (October 2016): 110-114.

Laland, K. N. "Social learning strategies." *Learning & Behavior* 32, no. 1 (2004): 4-14.

Lancy, D. F. "Homo Faber juvenalis: a multidisciplinary survey of children as tool makers/users." *Childhood Past* 10 (2017): 72-90.

Legare, C. H. et al. "Imitative flexibility and the development of cultural learning." *Cognition* 142 (September 2015): 351-361.

Legare, C. H., and M. Nielsen. "Imitation and Innovation: The Dual Engines of Cultural Learning." *Trends in Cognitive Sciences* 19, no. 11 (November 2015): 688-699.

Leont'ev, A. N. *Activity, Consciousness, and Personality.* Translated by M. J. Hall. Englewood Cliffs: Prentice-Hall, 1978.

Leont'ev, A. N. "The Psychological Principles of Preschool Play." In *The Development of the Mind: Selected Works of Aleksei Nikolaevich Leontyev*, edited by M. Cole, 331-354. Pacifica: Marxists Internet Archive, 2009.

Lewin, K. T. et al. "Level of aspiration." In *Personality and Behavior Disorders*, edited by J. M. Hunt, 333-378. Oxford: Ronald Press, 1944.

Lewontin, R. C. "Gene, organism and environment." In *Evolution from molecules to men*, edited by D. S. Bendall, 273-285. Cambridge: Cambridge University Press, 1983.

Liddell, T., and J. K. Kruschke. "Analyzing ordinal data with metric models: What could possibly go wrong?" *Open Science Framework*, April 2018.

Lyons, et al. " The scope and limits of overimitation in the transmission of artefact culture." *Philosophical Transactions of the Royal Society B* 366 (2011), 1158–1167.

Ma, L., and P. A. Ganea. "Dealing with conflicting information: Young children's reliance on what they see versus what they are told." *Developmental Science* 13, no. 1 (2004): 151-160.

Marsh, R. L., J. D. Landau, and J. L. Hicks. "How examples may (and may not) constrain creativity." *Memory & Cognition* 24, no. 5 (1996): 669-680.

McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* Boca Raton: CRC Press/Taylor & Francis, 2016.

McHugh, M. L. "Interrater reliability: the kappa statistic." *Biochemia Medica* 22, no. 3 (October 2012): 276-282.

Mecca, J. T., and M. D. Mumford. "Imitation and Creativity: Beneficial Effects of Propulsion Strategies and Specificity." *Journal of Creative Behavior* 48, no. 3 (2013): 209-236.

Meltzoff, A. N. "Understanding the Intentions of Others: Re-Enactment Intended Acts by 18-Month-Old Children." *Developmental Psychology* 31, no. 5 (September 1995): 838-850.

Merleau-Ponty, M. *Phenomenology of Perception.* Translated by C. Smith. London: Routledge & Kegan Paul, 1962.

Mesoudi, A., A. Thornton. What is cumulative cultural evolution? *Proceedings of the Royal Society B* 285 (2018): 20180712.

Mesoudi, A., A. Whiten, and R. Dunbar. "A bias for social information in human cultural transmission." *British Journal of Psychology* 97 (2006): 405-423.

Moreau, C. P., and M. G. Engeset. "The Downstream Consequences of Problem-Solving Mindsets: How Playing with Legos Influences Creativity." *Journal of Marketing Research* 53, no. 1 (February 2016): 18-30.

Morin, O. "Reasons to be fussy about cultural evolution." *Biology & Philosophy* 31, no. 3 (May 2016): 447-458.

Neldner, K., I. Mushin, and M. Nielsen. "Young children's tool innovation across culture: Affordance visibility matters." *Cognition* 168 (November 2017): 335-343.

Nielsen, M. et al. "Exploring tool innovation: a comparison of Western and Bushman children." *Journal of Experimental Child Psychology* 126 (2014): 384-394.

Nielsen, M. et al. "The persistent sampling bias in developmental psychology: a call to action." *Journal of Experimental Child Psychology* 162 (2017): 31-38.

Nielsen, M., and V. Slaughter. "Multiple Motivations for Imitation in Infancy." In *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social, and Communicative Dimensions*, edited by C. L. Nehaniv and K. Dautenhahn, 343-359. Cambridge: Cambridge University Press, 2007.

Nurmsoo, E., E. J. Robinson, and S. A. Butterfill. "Children's Selective Learning from Others." *Review of Philosophy and Psychology* 1, no. 4 (December 2010): 551-561.

Over, H., and M. Carpenter. "Putting the Social Into Social Learning: Explaining Both Selectivity and Fidelity in Children's Copying Behavior." *Journal of Comparative Psychology* 126, no. 2 (2012): 182-192.

Palagi, E., R. Stanyon, and E. Demuru. "Play to learn, teach by play." *Behavioral and Brain Sciences* 38 (2015): 35-36.

Peer, E. et al. "Beyond the Turk: an empirical comparison of alternative platforms for crowdsourcing behavioral research." *Journal of Experimental Social Psychology* 70 (May 2017): 153-163.

Pellegrini, A. D. *The role of play in human development.* New York: Oxford University Press, 2009.

Pellegrini, A. D., D. Dupuis, and P. K. Smith. "Play in evolution and development." *Developmental Review* 27, no. 2 (June 2007): 261-276.

Perreault, C., C. Moya, and R. Boyd. "A Bayesian approach to the evolution of social learning." *Evolution and Human Behavior* 33 (2012): 449-459.

Perry, S., B. J. Barrett, and I. Godoy. "Older, sociable white-faced capuchins (*Cebus capucinus*) invent more social behaviors; younger monkeys innovate more in other contexts." *PNAS USA* 114 (2017): 7806-7813.

Pinkham, A. M., and V. K. Jaswal. "Watch and learn? Infants privilege eficiency over pedagogy in imitative learning." *Infancy* 16, no. 5 (2011): 535-544.

Plotkin, H. C. "Learning and evolution." In *The role of behavior in evolution*, edited by H. C. Plotkin, 133-164. Cambridge: MIT Press, 1988.

Prolific Academic Ltd. *Home.* 2016. www.prolific.ac (accessed April 15, 2018).

Racine, J. S. "RStudio: A Platform-Independent IDE for R and Sweave." *Journal of Applied Econometrics* 27 (2012): 167-172.

Rakoczy, H., M. Tomasello, and T. Striano. "Young Children Know That Trying Is Not Pretending: A Test of the "Behaving-As-If" Construal of Children's Early Concept of Pretense." *Developmental Psychology* 40, no. 3 (May 2004): 388-399.

R Core Team. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing, 2015.

Rawlings, B., E. Flynn, and R. L. Kendal. "To copy or to innovate? The role of personality and social networks on children's learning strategies." *Child developmental perspectives* 11, no. 1 (2017): 39-44.

Reader, S. M., and K. N. Laland. "Animal Innovation: An Introduction." In *Animal Innovation*, edited by S. M. Reader and K. N. Laland, 3-35. Oxford: Oxford University Press, 2003.

Reader, S. M., and K. N. Laland. "Primate innovation: sex, age and social rank differences." *International Journal of Primatology* 22 (2001): 787-805.

Reindl, E. et al. "Young children copy cumulative technological design in the absence of action information." *Scientific Reports* 7 (2017): 1788.

Reindl, E., and C. Tennie. "Young children fail to generate an additive ratchet effect in an open-ended construction task." *PLoS ONE* 13, no. 6 (June 2018): e0197828.

Rendell, L. et al. "Cognitive culture: theoretical and empiricial insights into social learning strategies." *Trends in Cognitive Sciences* 15, no. 2 (February 2011): 68-76.

Rich, J. D., and R. W. Weisberg. "Creating All in the Family: A case study in creative thinking." *Creativity Research Journal* 16 (2004): 247-259.

Richerson, P. J., and R. Boyd. *Not By Genes Alone: How Culture Transformed Human Evolution.* Chicago: University of Chicago Press, 2005.

Riede, F. et al. "The role of play objects and object play in human cognitive evolution and innovation." *Evolutionary Anthropology* 27 (2018): 46-59.

Rogers, A. R. "Does Biology Constrain Culture?" *American Anthropologist* 90, no. 4 (1988): 819-831.

Rook, L. "Imitation in Creative Task Performance." *PhD Thesis.* Erasmus University Rotterdam, 2008.

Rook, L., and D. van Knippenberg. "Creativity and Imitation: Effects of Regularity Focus and Creative Exemplar Quality." *Creativity Research Journal* 23, no. 4 (2011): 346-356.

Rudman, H. et al. "Multidisciplinary exhibit design in a Science Centre: a participatory action research approach." *Educational Action Research*, August 2017.

Sanefuji, W. et al. "Emergence of the understanding of others' intentions: Re-enactment of intented acts from "failed attempts" in 12- to 24-year-olds." *Psychologia: An International Journal of Psychology in the Orient* 47 (2004): 10-17.

Searle, J. R. *Intentionality: An essay in philosophy of mind.* Cambridge: Cambridge University Press, 1983.

Seyfarth, R. M., and D. L. Cheney. "Precursors to language: Social cognition and pragmatic inferences in primates." *Psychonomic Bulletin & Review* 24, no. 1 (2016): 79-84.

Shalley, C. E., and J. E. Perry-Smith. "Effects of social-psychological factors on creative performance: The role of informational and controlling expected evaluation and modelling experience." *Organizational Behavior and Human Decision Processes* 84 (2001): 1-22.

Sheridan, K. M. et al. "The effects of environment and ownership on children's innovation of tools and tool material selection." *Philosophical Transactions of the Royal Society B* 371 (2016): 20150191.

Siegler, R. S. "Continuity and Change in the Field of Cognitive Development and in the Perspectives of One Cognitive Developmentalist." *Child Development Perspectives* 10, no. 2 (2016): 128-133.

Sinha, C. "Language and other artifacts: socio-cultural dynamics of niche construction." *Frontiers in Psychology* 6 (October 2015): 1601.

Smith, S. M., and S. E. Blankenship. "Incubation and the persistence of fixation in problem-solving." *American Journal of Psychology* 104 (1991): 61-87.

Smith, S. M., T. B. Ward, and J. S. Schumacher. "Constraining Effects of Examples in Creative Generation Task." *Memory & Cognition* 21, no. 6 (1993): 837-845.

Sterelny, K. *The Evolved Apprentice: How Evolution Made Humans Unique.* London: MIT Press, 2012.

Sternberg, R. J. *The Handbook of Creativity.* Cambridge: Cambridge University Press, 1999.

Stout, D. W. "Stone toolmaking and the evolution of human culture and cognition." *Philosophical Transactions of the Royal Society B* 366 (2011): 1050-1059.

Stout, D. W., and E. E. Hecht. "Evolutionary neuroscience of cumulative culture." *PNAS* 114, no. 30 (July 2017): 7861-7868.

Stuart Brown, L. *Play: How it Shapes the Brain, Opens the Imagination, and Invigorates the Soul.* New York: Penguin Group, 2010.

Subiaul, F. et al. "Imitation by cobination: Preschool age children evidence summative imitation in a novel problem-solving task." *Frontiers in Psychology* 6 (2015): 1-13.

Subiaul, F., E. M. Patterson, and R. Barr. "The cognitive strcuture of goal emulation during the preschool years." *The British Journal of Developmental Psychology* 34 (2016): 132-149.

Templeton, J. J. "Learning from others' mistakes: a paradox revisited." *Animal Behavior* 55 (February 1998): 79-85.

Tennie, C. et al. "Limitations to the cultural ratchet effect in young children." *Journal of Experimental Child Psychology* 126 (2014): 152-160.

Tennie, C. et al. "Ratcheting up the ratchet: On the evolution of cumulative culture." *Philosophical Transactions of the Royal Society B* 364 (2009): 2405-2415.

Thompson, D. E., and J. Russell. "The Ghost Condition: Imitation Versus Emulation in Young Children's Observational Learning." *Developmental Psychology* 40, no. 5 (2004): 882-889.

Tomasello, M. "Chimpanzee Social Cognition." *Monographs of the Society for Research in Child Development* 61, no. 3 (July 1996): 161-173.

Tomasello, M. "Cultural transmission in the tool use and communicatory signalling of chimpanzees?" In *"Language" and Intelligence in Monkeys and Apes*, 274-311. Cambridge: Cambridge University Press, 1990.

Truskanov, N., and Y. Prat. "Cultural transmission in an ever-changing world: trial-and-error copying may be more robust than precise imitation." *Philosophical Transactions of the Royal Society B* 373, no. 1743 (April 2018): 20170050.

Turner, C., L.-A. Giraldeau, and E. Flynn. "How does the reliability of a model affect children's choice to learn socially or individually." *Evolution and Human Behavior* 38 (2017): 341-349.

Užgiris, I. C. "Two functions of imitation during infancy." *International Journal of Behavioral Development* 4 (1981): 1-12.

Vale, G. L. et al. "Acquisition of a socially learned tool use sequence in chimpanzees: Implications for cumulative culture." *Evolution & Human Behavior* 38, no. 5 (September 2017a): 635-644.

Vale, G. L. et al. "Payoff-biased social learning." *Proceedings of the Royal Society B* 284, no. 1868 (December 2017b): 20171751.

van Oers, B. "Is it play? Towards a reconceptualisation of role play from an activity theory perspective." *European Early Childhood Education Research Journal* 21, no. 2 (2013): 185-198.

von Hofsten, C. et al. "Predictive action in infancy: Tracking and reaching for moving objects." *Cognition* 67 (1998): 255-285.

Vygotsky, L. S. "Play and its role in the mental development of the child." *Soviet Psychology* 5 (1967): 6-18.

Vygotsky, L. V. *Mind in Society.* London: Harvard University Press, 1978.

Walker, M. B., and M. G. Andrade. "Conformity in the Asche task as a function of age." *The Journal of Social Psychology* 136 (1996): 367-372.

Want, S. C., and P. L. Harris. "How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children." *Developmental Science* 5, no. 1 (2002): 1-41.

Want, S. C., and P. L. Harris. "Learning from Other People's Mistakes: Causal Understanding in Learning to Use a Tool." *Child Development* 72, no. 2 (March/April 2001): 431-443.

Weightman, G. *The Industrial Revolutionaries.* London: Atlantic Books, 2007.

Weisberg, D. S. et al. "Talking it up: play, language development, and the role of adult support." *American Journal of Play* 6 (2013): 39-54.

Weisberg, R. *Creativity: Genius and other myths.* New York: W. H. Freeman/Times Books/Henry Holt & Co., 1986.

Wellman, H. M., D. Cross, and J. Watson. "Meta-Analysis of Theory of Mind: The Truth about False Belief." *Child Development* 72, no. 3 (May/June 2001): 655-684.

Whalley, C. L., N. Cutting, and S. R. Beck. "The effect of prior experience on children's tool innovation." *Journal of Experimental Child Psychology* 161 (September 2017): 81-94.

Whitehead, H., and P. J. Richerson. "The evolution of conformist social learning can cause population collapse in realistically variable environments." *Evolution and Human Behavior* 30 (2009): 261-273.

Whitehouse, H. *Modes of religiosity: a cognitive theory of religious transmission.* Walnut Creek: AltaMira Press, 2004.

Whiten, A. et al. "How do apes ape?" *Learning & Behavior* 32 (2004): 36-52.

Whiten, A. et al. "Social learning in the real-world: 'Over-imitation' Occurs in Both Children and Adults Unaware of Participation in an Experiment and Independently of Social Interaction." *PLoS ONE* 11, no. 7 (2018): e0159920

Whiten, A., and E. Flynn. "The transmission and evolution of experimental microcultures in groups of young children." *Developmental Psychology* 46 (2010): 1694-1709.

Whiten, A., and E. van de Waal. "The pervasive role of social learning in primate lifetime development." *Behavioral Ecology and Sociobiology* 72 (2018): 80.

Whiten, A., and R. Ham. "On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research." In *Advances in the Study of Behavior*, edited by P. J. B. Slater, J. S. Rosenblatt, C. Beer and M. Milinski, 239-283. New York: Academic Press, 1992.

Williamson, R. A., A. N. Meltzoff, and E. M. Markman. "Prior experiences and perceived efficacy influence 3-year-old's imitation." *Developmental Psychology* 44, no. 1 (2008): 275-285.

Wilson, A. D., and S. Golonka. "Embodied cognition is not what you think it is." *Frontiers in Psychology* 4 (February 2013): Article 58.

Wood, L. A. N., R. L. Kendal, and E. G. Flynn. "Copy me or copy you? The effect of prior experience on social learning." *Cognition* 127 (February 2013a): 203-213.

Wood, L. A. N., R. L. Kendal, and E. G. Flynn. "Whom do children copy? Model-based biases in social learning." *Developmental Review* 33 (2013b): 341-356.

Woods, A. T., and et al. "Conducting perception research over the internet: a tutorial review." *PeerJ* 3 (2015): e1058.

Woodward, W. R. "The 'discovery' of social behaviorism and social learning theory, 1870-1980." *American Psychologist* 37, no. 4 (1982): 396-410.

Yamagishi, T. *Micro-Macro Dynamics of the Cultural Construction of Reality: A Niche Construction Approach to Culture.* Vol. 1, in *Advances in Culture and Psychology*, edited by M. J. Gelfand, C. Chiu and Y. Hong, 251-308. Oxford: Oxford University Press, 2011.

Zosh, J. M. et al. "Accessing the Inaccessible: Redefining Play as a Spectrum." *Frontiers in Psychology* 9 (August 2018): 1124.